Computational Molecular Biology

Lecture Seven: A Mathematical Diversion

More on longest common subsequences

Semester I, 2009-10

Graham Ellis
NUI Galway, Ireland

Permutations

There are 24 permutations of the four letters ACGT. These are:

AGCT	TGCA	TACG	TAGC
GACT	GTCA	ATCG	ATGC
GCAT	GCTA	ACTG	AGTC
CGAT	CGTA	CATG	GATC
CAGT	CTGA	CTAG	GTAC
ACGT	TCGA	TCAG	TGAC

Permutations

There are 24 permutations of the four letters ACGT. These are:

AGCT	TGCA	TACG	TAGC
GACT	GTCA	ATCG	ATGC
GCAT	GCTA	ACTG	AGTC
CGAT	CGTA	CATG	GATC
CAGT	CTGA	CTAG	GTAC
ACGT	TCGA	TCAG	TGAC

There are 20!=2432902008176640000 permutations of the 20 amino acids. One such permutation is

Phe-Ser-Tyr-Sys-Leu-Trp-Pro-Hys-Arg-Glu-Ile-Thr-Asn-Met-Lys-Val-Ala-Asp-Gly-Glu

Longest Common Subsequences of Permutations

The problem of finding longest common subsequences between permutations, say

$$\pi = \text{ACGT} \text{ and } \pi' = \text{CGAT},$$

can always be translated into a problem of finding longest common subsequences between permutations of integers, in this case

$$\tau = 1\ 2\ 3\ 4 \text{ and } \tau' = 2\ 3\ 1\ 4.$$

Longest Common Subsequences of Permutations

The problem of finding longest common subsequences between permutations, say

$$\pi = \text{ACGT} \text{ and } \pi' = \text{CGAT},$$

can always be translated into a problem of finding longest common subsequences between permutations of integers, in this case

$$\tau = 1\ 2\ 3\ 4 \text{ and } \tau' = 2\ 3\ 1\ 4.$$

The first letter of π' is the 2nd letter of π, the second leter of π' is the 3rd letter of π, the thrid letter of π' is the 1st letter of π, and the fourth letter of π' is the 4th letter of π.

Longest Increasing Subsequence of a Permutation

The problem of finding longest common subsequences between the integer permutations

$$\tau = 1\ 2\ 3\ 4 \text{ and } \tau' = 2\ 3\ 1\ 4.$$

is equivalent to the problem of finding a longest increasing subsequence of τ'.

Longest Increasing Subsequence of a Permutation

The problem of finding longest common subsequences between the integer permutations

$$\tau = 1\ 2\ 3\ 4 \text{ and } \tau' = 2\ 3\ 1\ 4.$$

is equivalent to the problem of finding a longest increasing subsequence of τ'.

For instance, 1 4 and 2 3 4 are both increasing subsequences of τ'.

Longest Increasing Subsequence of a Permutation

The problem of finding longest common subsequences between the integer permutations

$$\tau = 1\ 2\ 3\ 4 \text{ and } \tau' = 2\ 3\ 1\ 4.$$

is equivalent to the problem of finding a longest increasing subsequence of τ'.

For instance, 1 4 and 2 3 4 are both increasing subsequences of τ'.

The subsequence 3 1 4 is not increasing.

Problem

Find the length of a longest common subsequence between the permutations

$\pi =$ Phe-Ser-Tyr-Sys-Leu-Trp-Pro-Hys

and

$\pi' =$ Leu-Phe-Trp-Ser-Tyr-Pro-Sys-Hys

Problem

Find the length of a longest common subsequence between the permutations

$\pi =$ Phe-Ser-Tyr-Sys-Leu-Trp-Pro-Hys

and

$\pi' =$ Leu-Phe-Trp-Ser-Tyr-Pro-Sys-Hys

Solution

Just find a longest increasing subsequence of

$$\tau' = 5\ 1\ 6\ 2\ 3\ 7\ 4\ 8$$

Problem

Find the length of a longest common subsequence between the permutations

$\pi =$ Phe-Ser-Tyr-Sys-Leu-Trp-Pro-Hys

and

$\pi' =$ Leu-Phe-Trp-Ser-Tyr-Pro-Sys-Hys

Solution

Just find a longest increasing subsequence of

$$\tau' = 5\ 1\ 6\ 2\ 3\ 7\ 4\ 8$$

A longest increasing subsequence is 1 2 3 7 8 . So the length of a LCS is 5.

General Computational Problem

Given a permutation π of n integers, what algorithm can we use to find a longest increasing subsequence?

General Computational Problem

Given a permutation π of n integers, what algorithm can we use to find a longest increasing subsequence?

One possibility is to use our algorithm for finding longest common subsequences. (The special case of the Smith-Waterman algorithm.)

General Computational Problem

Given a permutation π of n integers, what algorithm can we use to find a longest increasing subsequence?

One possibility is to use our algorithm for finding longest common subsequences. (The special case of the Smith-Waterman algorithm.)

We could also use a very different algorithm due to Knuth, Robinson and Schensted.

General Computational Problem

Given a permutation π of n integers, what algorithm can we use to find a longest increasing subsequence?

One possibility is to use our algorithm for finding longest common subsequences. (The special case of the Smith-Waterman algorithm.)

We could also use a very different algorithm due to Knuth, Robinson and Schensted.

In order to explain this KRS algorithm we need to know about *partitions* and *Young tableaux*.

Partitions

A partition of a positive integer n is a sequence of integers

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$$

with

$$\lambda_1 + \lambda_2 + \cdots + \lambda_l = n.$$

Partitions

A partition of a positive integer n is a sequence of integers

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_l$$

with

$$\lambda_1 + \lambda_2 + \cdots + \lambda_l = n.$$

We write

$$(\lambda_1, \lambda_2, \ldots, \lambda_l) \vdash n.$$

Partitions of $n = 4$

$(4) \vdash 4$

$(3, 1) \vdash 4$

$(2, 2) \vdash 4$

$(2, 1, 1) \vdash 4$

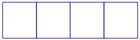
$(1, 1, 1, 1) \vdash 4$

Shapes of partitions

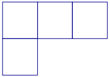
The shape of $\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_l) \vdash n$ is an array of l left-justified rows with each row i containing λ_i cells (or boxes).

Shapes of partitions of $n = 4$

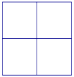
$(4) \vdash 4$



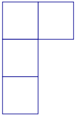
$(3, 1) \vdash 4$



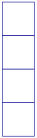
$(2, 2) \vdash 4$



$(2, 1, 1) \vdash 4$



$(1, 1, 1, 1) \vdash 4$



Young Tableaux

A Young tableau of shape $\lambda \vdash n$ is the shape with the numbers $1, 2, \cdots, n$ entered in the cells.

Young Tableaux

A Young tableau of shape $\lambda \vdash n$ is the shape with the numbers $1, 2, \cdots, n$ entered in the cells.

For example

1	3	4	7
2	5		
6			

is a Young tableau of shape $\lambda = (4, 2, 1) \vdash 7$.

Young Tableaux

A Young tableau of shape $\lambda \vdash n$ is the shape with the numbers $1, 2, \cdots, n$ entered in the cells.

For example

$$\begin{array}{|c|c|c|c|}\hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 \\ \cline{1-2} 6 \\ \cline{1-1}\end{array}$$

is a Young tableau of shape $\lambda = (4, 2, 1) \vdash 7$.

If each row is increasing from left to right, and each column is increasing from top to bottom, then the Young tableau is said to be a standard Young tableau.

Young Tableaux

A Young tableau of shape $\lambda \vdash n$ is the shape with the numbers $1, 2, \cdots, n$ entered in the cells.

For example

1	3	4	7
2	5		
6			

is a Young tableau of shape $\lambda = (4, 2, 1) \vdash 7$.

If each row is increasing from left to right, and each column is increasing from top to bottom, then the Young tableau is said to be a standard Young tableau.

The above example is a standard Yound tableau.