

On the complexity of SNP Block Partitioning Under the Perfect Phylogeny Model

Jens Gramm¹ Tzvika Hartman² Till Nierhoff³
oded sharan⁴ **Till Tantau**⁵

¹Universität Tübingen, Germany

²Bar-Ilan University, Ramat-Gan, Israel

³International Computer Science Institute, Berkeley, USA

⁴Tel-Aviv University, Israel

⁵Universität zu Lubeck, Germany

OUTLINE

1 Introduction

- The Model and the Problem
- The Integrated Approach

2 Bad News : Hardness Results

- Hardness of PP-Partitioning of Haplotype Matrices
- Hardness of PP-Partitioning of Genotype Matrices

3 Good News : Tractability Results

- Perfect Path Phylogenies
- Tractability of PPP-Ppartitioning of Genotype Matrices

General formalisation of haplotyping

INPUTS

- A **genotype matrix** G .
- The **rows** of the matrix are **taxa / individuals**.
- The **columns** of the matrix are **SNP sites / characters**.

OUTPUTS

- A **haplotype matrix** H .
- Pairs of rows in H **explain** the rows of G
- The haplotypes in H are **biologically plausible**.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

We can do perfect phylogeny haplotyping efficiently, but ...

1 Data may be missing.

- This makes the problem NP-complete ...
- ...even for very restricted cases

Solutions :

- Additional assumption like the rich data hypothesis.

2 No perfect phylogeny is possible.

- This can be caused by chromosomal crossing-over effects.
- This can be caused by incorrect data.
- This can be caused by multiple mutations at the same sites.

Solutions :

- Look for phylogenetic networks.
- Correct for

3 Find bloks where a perfect phylogeny is possible

Vibrations longitudinales

célérité des ondes longitudinales

- $[c] = \text{m/s}$
- c = célérité du son ou vitesse des ondes longitudinales
- Ordres de grandeur :

$$c_{\text{acier}} = \sqrt{\frac{E}{\delta}} = \frac{2 \cdot 10^{11}}{8 \cdot 10^3} = 5 \cdot 10^4 = 5000 \text{ m/s}$$

Matériau	c(m/s)	Matériau	c(m/s)
PVC mou	80	Glace	3200
Sable sec	10-300	Hêtre	3300
Béton	3100	Aluminium	5035
Plomb	1200	Verre	5300
PVC dur	1700	Acier	5600-5900
Granit	6200	Péridotite	7700

Rappel : vitesse du son dans l'air : 343 m/s, dans l'eau : 1480 m/s

Lipsum 6

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

OUTLINE

1 Introduction

- The Model and the Problem
- The Integrated Approach

2 Bad News : Hardness Results

- Hardness of PP-Partitioning of Haplotype Matrices
- Hardness of PP-Partitioning of Genotype Matrices

3 Good News : Tractability Results

- Perfect Path Phylogenies
- Tractability of PPP-Ppartitioning of Genotype Matrices

Lipsum 7

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Lipsum 8

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

OUTLINE

1 Introduction

- The Model and the Problem
- The Integrated Approach

2 Bad News : Hardness Results

- Hardness of PP-Partitioning of Haplotype Matrices
- Hardness of PP-Partitioning of Genotype Matrices

3 Good News : Tractability Results

- Perfect Path Phylogenies
- Tractability of PPP-Ppartitioning of Genotype Matrices

Lipsum 9

Morbi luctus, wisi viverra faucibus pretium, nibh est placerat odio, nec commodo wisi enim eget quam. Quisque libero justo, consectetur a, feugiat vitae, porttitor eu, libero. Suspendisse sed mauris vitae elit sollicitudin malesuada. Maecenas ultricies eros sit amet ante. Ut venenatis velit. Maecenas sed mi eget dui varius euismod. Phasellus aliquet volutpat odio. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Pellentesque sit amet pede ac sem eleifend consectetur. Nullam elementum, urna vel imperdiet sodales, elit ipsum pharetra ligula, ac pretium ante justo a nulla. Curabitur tristique arcu eu metus. Vestibulum lectus. Proin mauris. Proin eu nunc eu urna hendrerit faucibus. Aliquam auctor, pede consequat laoreet varius, eros tellus scelerisque quam, pellentesque hendrerit ipsum dolor sed augue. Nulla nec lacus.

Lipsum 10

Suspendisse vitae elit. Aliquam arcu neque, ornare in, ullamcorper quis, commodo eu, libero. Fusce sagittis erat at erat tristique mollis. Maecenas sapien libero, molestie et, lobortis in, sodales eget, dui. Morbi ultrices rutrum lorem. Nam elementum ullamcorper leo. Morbi dui. Aliquam sagittis. Nunc placerat. Pellentesque tristique sodales est. Maecenas imperdiet lacinia velit. Cras non urna. Morbi eros pede, suscipit ac, varius vel, egestas non, eros. Praesent malesuada, diam id pretium elementum, eros sem dictum tortor, vel consectetur odio sem sed wisi.