

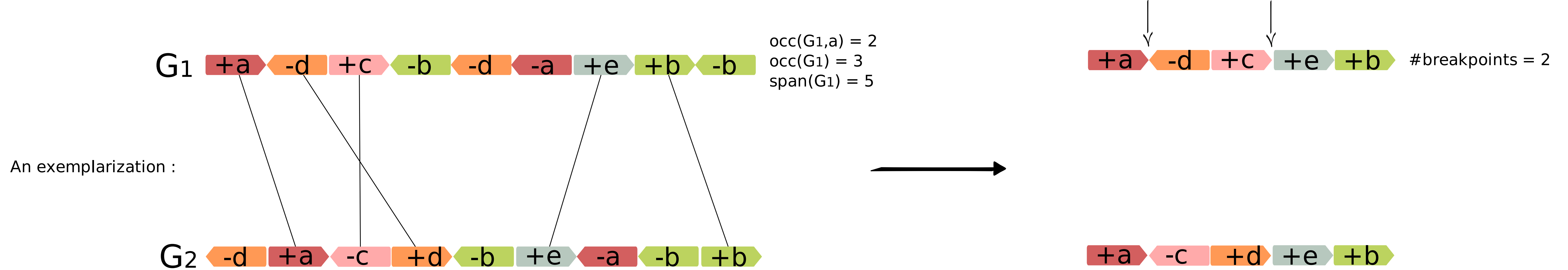
# The EXEMPLAR BREAKPOINT DISTANCE is not approximable... at all

G. Blin<sup>1</sup>, G. Fertin<sup>2</sup>, F. Sikora<sup>1</sup>, S. Vialette<sup>1</sup>

<sup>1</sup> Université Paris-Est, IGM-LabInfo - UMR CNRS 8049 - France  
<sup>2</sup> LINA - UMR CNRS 6241 - Université de Nantes - France

**Abstract :** A promising and active field of comparative genomics consists in comparing two genomes by establishing a one-to-one correspondence (i.e., a matching) between their genes. This correspondence is usually chosen in order to optimize a predefined measure. One such problem is the Exemplar Breakpoint Distance problem (or EBD, for short), which asks, given two genomes modeled by signed sequences of characters, to keep and match exactly one occurrence of each character in the two genomes (a process called exemplarization), so as to minimize the number of breakpoints of the resulting genomes. The gene which is kept is assumed to be the ancestral gene. Bryant [3] showed that EBD is NP-complete. In this paper, we close the study of the approximation of EBD by showing that no approximation factor can be derived for EBD considering non-trivial genomes – i.e. genomes that contain duplicated genes.

## Notations and problems



### The Exemplar Breakpoint Distance (EBD) Problem :

**Input :** two genomes **G1**, **G2** and an integer **k**  
**Question :** is it possible to establish an exemplar matching of **G1** and **G2**, such that the breakpoint distance between the resulting genomes is at most **k** ?

### The Zero Exemplar Breakpoint Distance (ZEBD) Problem :

**Input :** two genomes **G1**, **G2**  
**Question :** is it possible to establish an exemplar matching of **G1** and **G2**, such that the breakpoint distance between the resulting genomes is zero ?

## Known results

For any  $p, q \geq 1$ , we denote EBD( $p, q$ ) (resp. ZEBD( $p, q$ )) the EBD (resp. ZEBD) problem in which  $\text{occ}(G1)=p$  and  $\text{occ}(G2)=q$ .

### On Exemplar Breakpoint Distance (EBD) :

- EBD(1,2) is NP-Complete [3]
- EBD(1,2) is APX-Hard [2]
- EBD(3,3) got no approximation algorithm [4]

**Open problem :** is EBD(2,2) approximable ? [2,4]

### On Zero Exemplar Breakpoint Distance (ZEBD, EBD with $k=0$ ) :

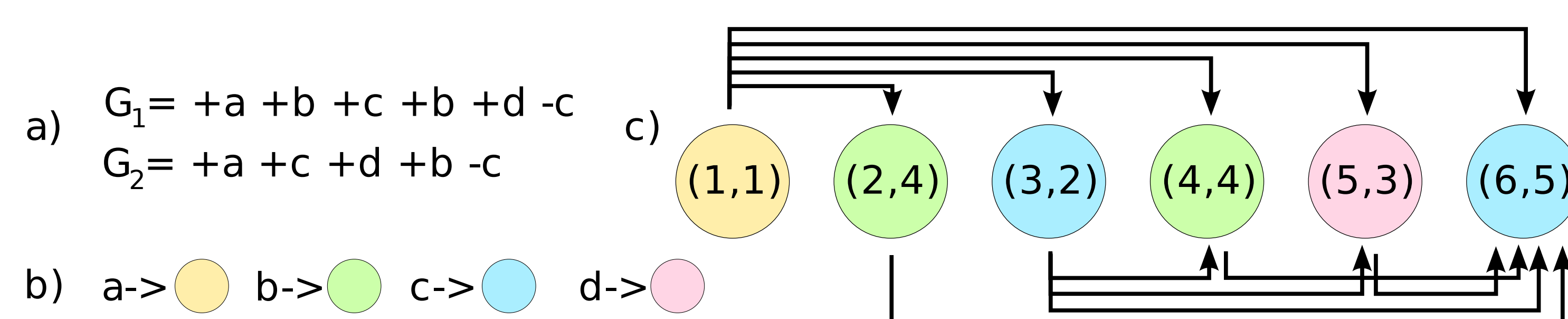
- ZEBD(3,3) is NP-Complete [4]
- ZEBD(2,q) is NP-Complete (q unbounded) [2]

**Open problem :** is ZEBD(2,q) NP-Complete for bounded q ?

## Our contributions

- ZEBD(2,2) is NP-Complete (reduction from 3-SAT), and hence EBD for non trivial genomes is not approximable at all.
- Two Fixed Parameter Tractable (FPT) algorithms for ZEBD :

1) Parameterized by the number **m** of genes families



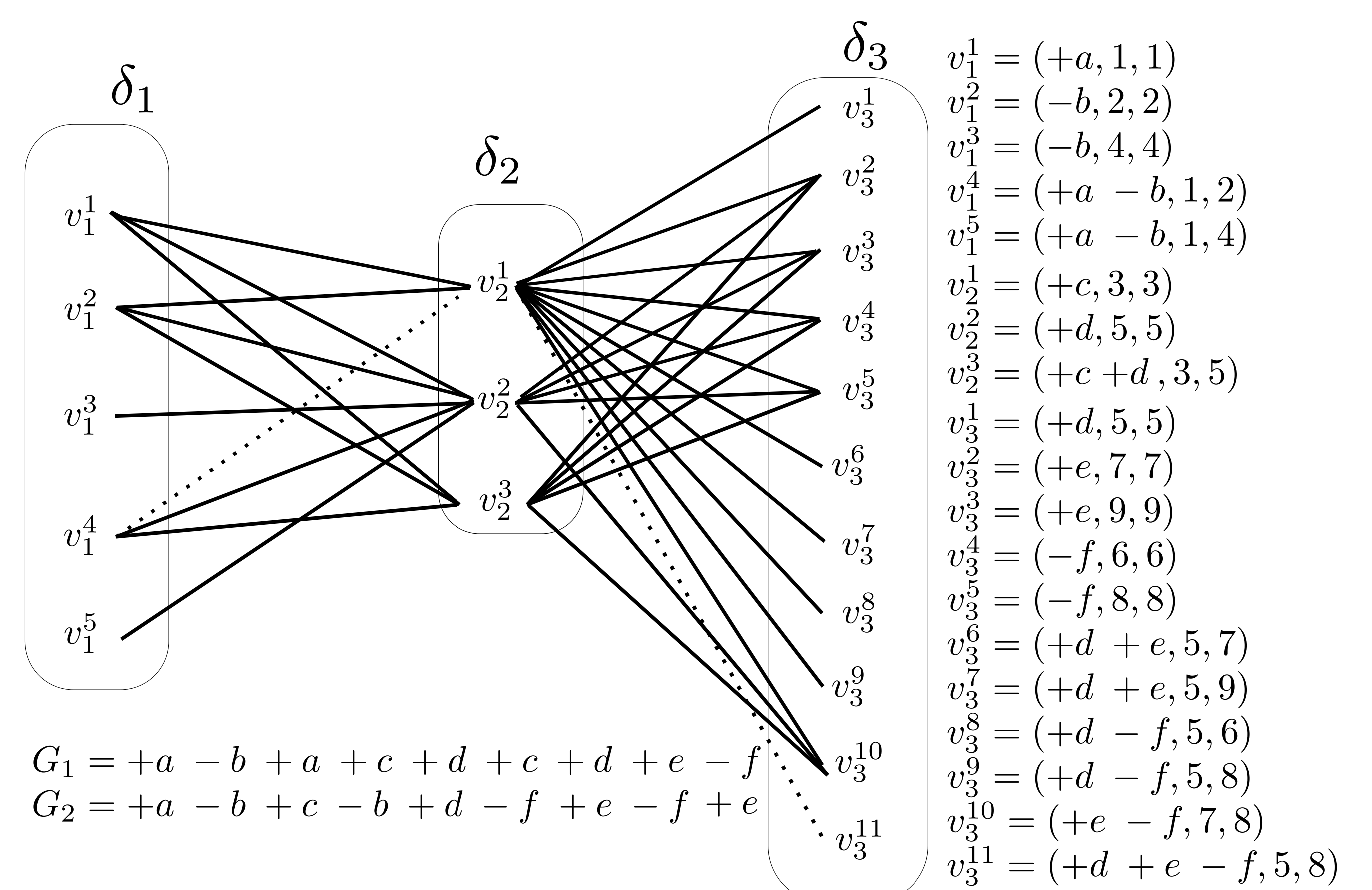
By color-coding [1]

- A color for each gene family
- A vertex  $v \in V$  for each pair of genes of the same family that carry the same sign
- For all  $\{(i,j), (p,q)\} \in V^2$ , an edge from (i,j) to (p,q) if  $i < p$  and  $j < q$

Looking for a colorful path

**Complexity :**  $O(m2^m)$

2) Parameterized by genome's span **s**



- A vertex : a common subsequence between  $G_1$  and  $G_2$
- An edge if the subsequences are compatible (no gene of the same family in common and not overlapping)

Compute the common subsequences between segments of  $G_1$  of size  $s$  and  $G_2$

Looking for a path in the DAG

**Complexity :**  $O(n2^s s^3)$

## References

1. N. Alon, R. Yuster, and U. Zwick. Color coding. Journal of the ACM, 42(4):844–856, 1995.
2. S. Angibaud, G. Fertin, I. Rusu, A. Thévenin, and S. Vialette. On the approximability of comparing genomes with duplicates. Journal of Graph Algorithms and Applications, 2008.
3. D. Bryant. The complexity of calculating exemplar distances. In Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment, and the Evolution of Gene Families, pages 207–212. Kluwer Academic Publisher, 2000.
4. Z. Chen, B. Fu, and B. Zhu. The approximability of the exemplar breakpoint distance problem. In 2nd International Conference on Algorithmic Aspects in Information and Management (AAIM), volume 4041 of Lecture Notes in Computer Science, pages 291–302. Springer, 2006.