> Nantes
2—5
juillet

JOBIM 2019

JOURNÉES OUVERTES DE BIOLOGIE INFORMATIQUE & MATHÉMATIQUES

Thèmatiques :

Biologie structurale
Biologie des systèmes
Epidémiologie Génétique
Evolution/Phylogénie
Génomique/Métagénomique
Sciences des données

Abstracts

Keynotes :

Chloé-Agathe Azencott, Paris
Alexander Bockmayr, Berlin
Alessandra Carbone, Paris
Olivier Delaneau, Lausanne
Christophe Dessimoz, Lausanne
Juliette Martin, Lyon

https://jobim2019.sciencesconf.org

Dear attendees of the 20th edition of JOBIM, welcome in Nantes !

JOBIM is the French national conference dedicated to promoting an active interface between Biology, Computer Sciences, and Mathematics. After a previous visit to Nantes in 2009, JOBIM comes back this year in this same city from western France. Since the last visit, the bioinformatics community has impressively grown, and new fields are today covered. The impressive number of submissions at JOBIM 2019 reflects such an increase in our community. The Program Committee received a total of 260 submissions deciphered as 29 long presentations, 11 flash presentations, 15 demos and 204 posters. As a main novelty of the 2019th edition, JOBIM 2019 will present five additional thematic sessions that will cover particular topics more specialized.

We sincerely thank all the members of the Program Committee who helped us to set up a great scientific program by reviewing all submissions in time. This task would have been impossible without them! We also are grateful to the six invited speakers that have accepted to contribute to the success of the JOBIM edition in Nantes.

We are indebted to the organizing institutions, the SFBI, the GDR BIM, and the IFB. We are also grateful to all our partners and sponsors for their financial support.

Finally, we could also not forget to warmly thank Sophie Girault, Elodie Guidon, Aurore Morvan, and Jérémie Ségard as well as all the members of the organizing committee who worked collectively without counting sweat and tears to welcome the cream of bioinformaticians today in the best conditions.

Damien Eveillard and Audrey Bihouée
Organizing committee

Jérémie Bourdon and Richard Redon
Program committee

Our Partners and Sponsors

Contents

A web server for identification and analysis of coevolution in overlapping proteins

Elin Teppa[1], Alessandra Carbone[1,2]

1 Laboratory of Computational and Quantitative Biology (LCQB), Sorbonne Université, CNRS, IBPS, UMR7238, 4, place Jussieu 75005 Paris, France.
2 Institut Universitaire de France (IUF) 75005 Paris, France

Corresponding author:elinteppa@gmail.com

Abstract

Overlapping genes exist in all domains of life and are especially abundant in viral genomes. The existence of overlapping reading frames increases the rising of deleterious mutations for one of the proteins, since a single nucleotide substitution may affect both proteins. Molecular coevolution may be seen as a mechanism to tolerate or compensate unfavorable mutations, decreasing the evolutionary constraints in the overlapping region. For instance, a favorable mutation in one reading frame may be unfavorable in the other reading frame and additional mutations may be needed to compensate the first mutation. Although molecular coevolution was widely used in viral genomes, the "overlap problem" was disregarded. Here, we present a server that facilitates the analysis of coevolution in overlapping proteins and of the impact of mutations in another ORF.

Keywords: coevolution; compensatory mutations, virus, overlapping proteins

Introduction

Multiple studies of coevolving positions in viral sequences have been useful to understand functionally significant residues [1,2], to predict protein-protein interactions [3], to modulate viral fusion [4] and to identify drug resistance mutations [5–9] among others.

The genomes of most viral species have overlapping genes—two or more proteins coded for by the same nucleotide sequence. ORFs may overlap in various manners considering the type, the direction of transcription and the ORFs' phase (Fig 1). Sequence analysis in overlapping ORFs represents a challenge due to changes in the nucleotide sequence that may simultaneously affect both proteins within their overlapping region.

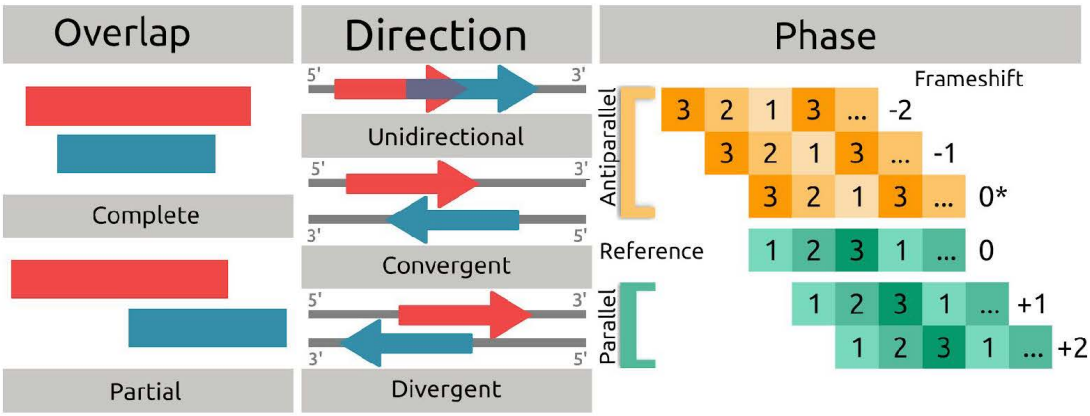

Figure 1: Definitions on ORFs overlap.

An overlap between two ORFs can be complete (if an ORF is nested within the other) or partial (if only the 3' or 5' end are overlapping). ORFs can overlap on the same strand, or in the case of a double-stranded genome, on the reverse complementary strand. Hence, three directions are possibles: unidirectional, convergent and divergent. The reference ORF, in a pair of overlapping ORFs, is called phase 0. Overlaps in a parallel strand can be in two phases whereas antiparallel-strand overlaps can be in three phases.

Given that coevolution may be seen as a mechanism to tolerate or compensate unfavorable mutations, molecular coevolution in the overlapping region may help to decrease evolutionary constraints. As far as we know, there is no study of coevolution that considers both overlapped proteins.

In the overlapping region, coevolution in an ORF: may be mirrored by coevolution in the other ORF; may generate a non-synonymous substitution which in turn may be compensated by other mutations (inside or outside the overlapping region); may generate synonymous substitutions (Fig 2).

The motivation for this server is to provide a tool to facilitate the analysis of coevolution in overlapped protein and of the impact of mutations in another ORF. To do that we combine information at protein and nucleotide levels.

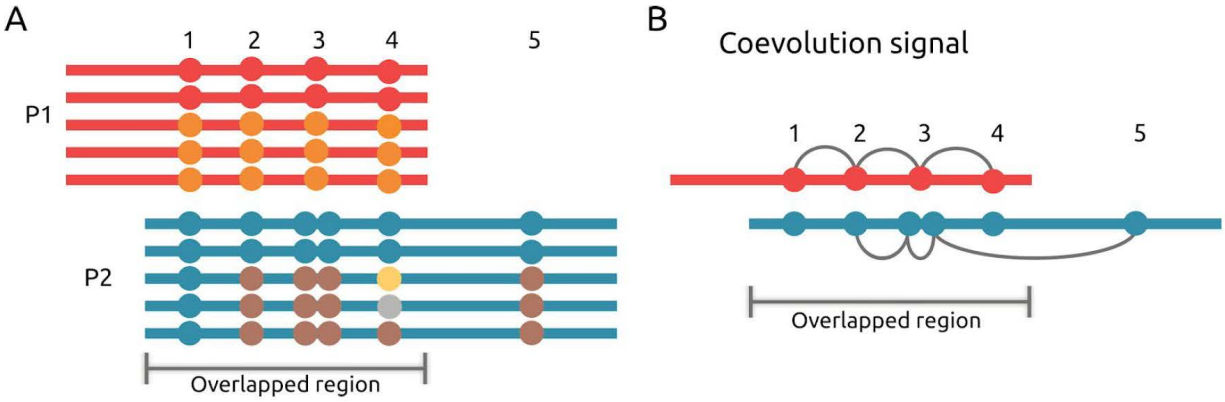

Figure 2: Coevolution pattern in overlapping region.

Different effects of four coevolving positions in the overlapped region of two proteins (P1 and P2). **A:** A cluster of four coevolving positions is represented in P1's alignment where two sequences maintain the wild-type residues (red circles) and three sequences show mutations on all positions (orange circles). A

mutation in P1 may be coupled by synonymous substitutions in P2 (column 1); the same non-synonymous substitution (column 2), two non-synonymous substitutions in adjacent positions (column 3); a variety of non-synonymous substitutions (column 4). The cluster of coevolving positions may also contain positions outside the overlapping region (column 5). **B:** P1 shows a coevolution signal between the first four positions (gray lines) which partially coincides with coevolution detected in P2.

Methods

Input

The input is a nucleotide alignment of the pair of overlapping protein sequences to be analyzed. It will contain the overlapped and non-overlapped regions of both proteins, as well as the start and end positions of the proteins and their corresponding DNA strand (parallel or antiparallel) (Figure 1).

Workflow

Given a DNA alignment and its associated distance tree that can be provided or optionally generated automatically, all subsets of sequences corresponding to the subtrees of the tree are systematically considered for coevolution analysis. For each subset, the ORF1 and ORF2 sequences (Fig 2) are translated into amino acids and the resulting protein alignments are used as input to predict coevolving positions using the BIS2 algorithm [10,11]. Our iterative strategy allows applying BIS2 in a large number of conserved sequences. As part of the result, the clusters of coevolving positions detected for both proteins are provided. If coevolution is detected in the overlapping region for one of the proteins, the effect of variation is analyzed in the other protein. By analyzing the subset of sequences where the cluster is detected for the first protein, we identify if the coevolving positions are accompanied by one or more synonymous/non-synonymous substitution(s) and if these positions also show coevolution in the second protein.

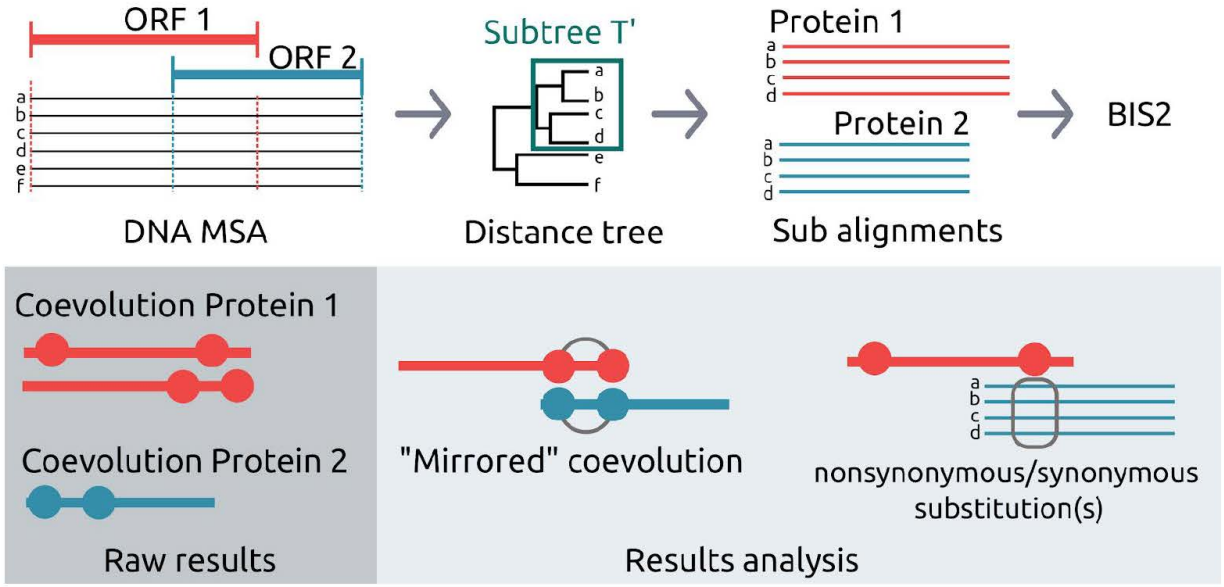

Figure 3: Schematic representation of the workflow from the input sequences to the results.

The DNA alignment covering both ORFs to be analyzed is used to generate a distance tree, optionally the tree may be provided by the user. Then, the tree is partitioned in all possible subtrees. The protein sequences corresponding to the subtrees are used as input to compute coevolution using BIS2 algorithm. The results include the coevolution of each of the proteins, as well as the effect of the mutations of one protein on the other. It is also indicated if both proteins show coevolution in equivalent positions ("mirrored" coevolution) or if the mutation of the co-evolved position in a protein is accompanied by synonymous or nonsynonymous mutations in the other.

Conclusions

We have developed an interactive web server providing an intuitive representation of of the coevolved residues predicted in overlapping proteins. To the best of our knowledge, this is the only publicly available method designed to analyze coevolution in overlapping protein sequences. The server is simple to use and it provides a powerful tool the virologist and the biologist to compute coevolution and analyze the effect of mutations in overlapping regions. Its results should help to elucidate the evolutionary constraints found in overlapping ORFs.

Acknowledgements

This work was supported by the French "Agence Nationale de la Recherche sur le SIDA et les hépatites virales" (ANRS CSS4 ECTZ25224 – 2017-19 to AC; www.anrs.fr).

References

1. Le L, Leluk J. Study on phylogenetic relationships, variability, and correlated mutations in M2 proteins of influenza virus A. PLoS One. 2011;6: e22970.
2. Jain J, Mathur K, Shrinet J, Bhatnagar RK, Sunil S. Analysis of coevolution in nonstructural proteins of chikungunya virus. Virol J. 2016;13: 86.
3. Champeimont R, Laine E, Hu S-W, Penin F, Carbone A. Coevolution analysis of Hepatitis C virus genome to identify the structural and functional dependency network of viral proteins. Sci Rep. 2016;6: 26401.
4. Douam F, Fusil F, Enguehard M, Dib L, Nadalin F, Schwaller L, et al. A protein coevolution method uncovers critical features of the Hepatitis C Virus fusion mechanism. PLoS Pathog. 2018;14: e1006908.
5. Rhee S-Y, Liu TF, Holmes SP, Shafer RW. HIV-1 subtype B protease and reverse transcriptase amino acid covariation. PLoS Comput Biol. 2007;3: e87.
6. Handel A, Regoes RR, Antia R. The role of compensatory mutations in the emergence of drug resistance. PLoS Comput Biol. 2006;2: e137.
7. González-Ortega E, Ballana E, Badia R, Clotet B, Esté JA. Compensatory mutations rescue the virus replicative capacity of VIRIP-resistant HIV-1. Antiviral Res. 2011;92: 479–483.
8. Bloom JD, Gong LI, Baltimore D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science. 2010;328: 1272–1275.
9. Tanaka MM, Valckenborgh F. Escaping an evolutionary lobster trap: drug resistance and compensatory mutation in a fluctuating environment. Evolution. 2011;65: 1376–1387.
10. Dib L, Carbone A. Protein fragments: functional and structural roles of their coevolution networks. PLoS One. 2012;7: e48124.
11. Oteri F, Nadalin F, Champeimont R, Carbone A. BIS2Analyzer: a server for co-evolution analysis of conserved protein families. Nucleic Acids Res. 2017;45: W307–W314.