

CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes

Wolf T, Shelest V, Nath N, Shelest E (2015) CASSIS and SMIPS: Promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes *Bioinformatics* 32(8), 1138-1143.

[Details](#)



Abstract

MOTIVATION: Secondary metabolites (SM) are structurally diverse natural products of high pharmaceutical importance. Genes involved in their biosynthesis are often organized in clusters, i.e., are co-localized and co-expressed. In silico cluster prediction in eukaryotic genomes remains problematic mainly due to the high variability of the clusters' content and lack of other distinguishing sequence features. **RESULTS:** We present Cluster Assignment by Islands of Sites (CASSIS), a method for SM cluster prediction in eukaryotic genomes, and Secondary Metabolites by InterProScan (SMIPS), a tool for genome-wide detection of SM key enzymes ('anchor' genes): polyketide synthases, non-ribosomal peptide synthetases and dimethylallyl tryptophan synthases. Unlike other tools based on protein similarity, CASSIS exploits the idea of co-regulation of the cluster genes, which assumes the existence of common regulatory patterns in the cluster promoters. The method searches for 'islands' of enriched cluster-specific motifs in the vicinity of

anchor genes. It was validated in a series of cross-validation experiments and showed high sensitivity and specificity. AVAILABILITY AND IMPLEMENTATION: CASSIS and SMIPS are freely available at <https://sbi.hki-jena.de/cassis>. CONTACT: thomas.wolf@leibniz-hki.de or ekaterina.shelest@leibniz-hki.deSupplementary information: Supplementary data are available at Bioinformatics online.

Involved units

[Microbiome Dynamics](#) [Gianni Panagiotou](#) [Read more](#)

Leibniz-HKI-Authors



Neetika Nath

[Details](#)



Ekaterina Shelest

[Details](#)



Vladimir Shelest

[Details](#)



Thomas Wolf

[Details](#)

Topics

[Prediction of gene regulatory elements in fungi](#)

Identifier

doi: 10.1093/bioinformatics/btv713

PMID: 26656005