

Laura Pellegrini
Matricola: 592743

RELAZIONE DI BIOLOGIA MOLECOLARE:

Argomento scelto: Descrivi il contenuto del Database JASPAR e l'utilità delle matrici di peso che contiene

1- IN CHE COSA CONSISTE IL DB JASPAR?

Jaspar è un database contenente fattori di trascrizione di alta qualità che preferenzialmente legano il DNA. I dati si presentano sotto forma di matrici e possono essere facilmente convertiti in Position Weight Matrices (PSSM).

Jaspar più precisamente è una collezione di database più piccoli, i quali hanno obiettivi differenti.

La maggior parte dei ricercatori utilizza JASPAR CORE.

Questo database contiene set di profili (PSSM) curati, non ridondanti, ottenuti da pubblicazioni scientifiche. Tutti i profili sono derivati da studi sui siti di legame per i fattori di trascrizione in eucarioti pluricellulari.

Quindi il database rappresenta un'accurata collezione di sequenze target.

Uno dei principali obiettivi di JASPAR CORE è quello di assegnare un unico modello per ogni fattore di trascrizione: ciò è importante nella ricerca perché porta ad avere un database non ridondante.

Per convenienza, JASPAR CORE è ulteriormente suddiviso in sottogruppi però questa distinzione è utilizzata solo nell'interfaccia web.

The high-quality transcription factor binding profile database

Browse the JASPAR CORE database directly:

- JASPAR CORE Vertebrata
- JASPAR CORE Nematoda
- JASPAR CORE Insecta
- JASPAR CORE Plantae
- JASPAR CORE Fungi
- JASPAR CORE by Structural Class

DOCUMENTATION DOWNLOAD CONTACT

Browse/search a JASPAR database

Select a JASPAR database ?

JASPAR CORE

JASPAR Collections

- JASPAR CNE
- JASPAR FAM
- JASPAR PBM
- JASPAR PBM_HLH
- JASPAR PBM_HOMEO

The JASPAR CORE database contains a curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes. The prime difference to similar resources (TRANSFAC, etc) consist of the open data access, non-redundancy and quality.

When should it be used? When seeking models for specific factors or structural classes, or if experimental evidence is paramount

Browse sorted by ?

- ID
- Species
- Structural class
- Taxonomic group

[Browse](#)

Search by ?

Name [AND](#)

Name [AND](#)

Name [Search](#)

Align to a custom matrix or IUPAC string ?

[Azzera](#) | [Fill in an example matrix](#) | [Align](#)

2- QUALI SONO I CAMPI DEL DATABASE JASPAR CORE?

- ID : un unico identificatore per ogni modello
- Name: il nome del fattore di trascrizione. Quando è possibile il nome è basato sulla standardizzazione in simboli di Entrez gene.
- Class: classe strutturale del fattore di trascrizione, basata sul sistema TFCaT.
- Family: sottoclasse strutturale del fattore di trascrizione, basata sul sistema TFCaT
- Species: la species source per la sequenza, in Latino, presa dalla NCBI Taxonomic browser.
- Tax_group: il gruppo di specie
- Acc: l'Accession number rappresentativo per il fattore di trascrizione in Gene Bank
- Type: la metodologia usata per costruire le matrici
- Medline: un link a una pubblicazione rilevante che riporta il sito usato nella costruzione nel modello
- Comment

QUANDO DOVREBBE ESSERE USATO QUESTO DATABASE?

Quando si ricercano modelli per fattori specifici o classi strutturali, o quando l'evidenza sperimentale è di primaria importanza.

JASPAR COLLECTION:

Jaspar collections è una collezione di matrici che sono utili ma non rientrano nell'ambito di JASPAR CORE.

1- JASPAR FAM:

Questo database consiste in 11 modelli che descrivono proprietà di legame condivise di classi strutturali di fattori di trascrizione. Questi modelli possono essere chiamati matrici di consenso. Essi hanno 2 vantaggi principali: 1) riduzione della complessità dei risultati, 2) i modelli possono essere usati per classificare profili nuovi.

Questo database andrebbe usato nella ricerca di grandi sequenze genomiche di cui non si hanno precedenti informazioni. Oppure per la classificazione di nuovi profili forniti dall'utente.

2- JASPAR PHYLOFACTS:

Questo database consiste in 174 profili estratti da sequenze geniche filogeneticamente conservate. In breve, i promotori di geni umani (presi dal database RefSeq) sono stati allineati con i genomi di topo, ratto e cane. Viene in seguito effettuata la scansione di ogni sequenza consenso di lunghezza compresa tra 6 e 26 aa lungo l'allineamento. Un motivo è da considerarsi conservato se è presente nell'allineamento sia per l'uomo, sia per le altre 3 specie.

Il rapporto di conservazione "p" è definito come il numero di volte in cui un motivo è conservato diviso il numero di volte in cui è presente solo nell'uomo.

Questo rapporto di conservazione è confrontato con il rapporto di conservazione atteso p_0 , stimato a partire da motivi casuali, per valutare l'affidabilità del metodo.

Le matrici di JASPAR PHYLOFACTS sono un mix di motivi per fattori di trascrizione noti e non definiti. Sono utili quando ci si aspetta che altri fattori possono determinare caratteristiche dei promotori, come per esempio fattori strutturali o specificità tissutale.

3- JASPAR POLII:

L'enormità di nuovi dati sperimentali riguardanti i siti di inizio della trascrizione ha motivato lo studio computazionale dei core promoters.

Il sub-database JASPAR POLII contiene 13 pattern noti di DNA legati ai promotori del core della RNA polimerasi II.

Dovrebbe essere usato per l'analisi di promotori del core.

4- JASPAR CNE:

Jaspar CNE è una collezione di 233 matrici di profilo contenenti elementi non codificanti altamente conservati, enhancer, e altri elementi coinvolti nella regolazione dell'espressione genica.

Viene appunto utilizzato per l'analisi di potenziali enhancer.

5- JASPAR SPLICE:

Jaspar SPLICE è una piccola collezione di matrici di profilo per siti di splicing canonici e non canonici presenti nel genoma umano. Per ora contiene solo 6 profili ottenuti dal genoma umano. Si dovrebbe usare per analizzare siti di splicing e per lo studio dello splicing alternativo.

6- JASPAR PBM e JASPAR PBM HOMEIO:

Tutte le collezioni PBM sono state costruite con l'utilizzo di nuove tecniche in vitro basate sui microarrays. I modelli di matrici PBM hanno il loro database specifico: UniPROBE .

Entrambe le collezioni contengono profili derivanti da fattori di trascrizione e homeodomains di topo.

7- JASPAR PBM HLH:

Anche questa è una collezione PBM costruita mediante nuove tecniche in vitro.

PBM HLH contiene 19 modelli bHLH (basic helix loop elix, motivo strutturale che caratterizza i fattori di trascrizione) studiati nell'organismo modello *C. elegans*.

3- QUAL E' L'UTILITA' DELLE MATRICI DI PESO CHE CONTIENE JASPAR?

Le matrici di peso o profili, sono una forma di punteggio dipendente dalla posizione (Position Specific Scoring Matrix). Esse permettono di usare l'informazione che si ottiene dall'allineamento progressivo, come per esempio il grado di conservazione di un residuo in una certa posizione, e permettono inoltre di determinare la sequenza consenso.

L'utilità di questi profili sta nel fatto che ci forniscono informazioni sia di tipo qualitativo che di tipo quantitativo. Essi derivano dalla sintesi di matrici di frequenza, che a partire da un allineamento multiplo calcolano la frequenza di un amminoacido in quella posizione dell'allineamento, e matrici di sostituzione, le quali attribuiscono un determinato punteggio alle sostituzioni amminoacidiche in base alla frequenza della sostituzione.

L'idea di creare delle matrici di peso è nata dai grafici chiamati "sequence logos", presenti anche nel database Jasper, che rappresentano la frequenza dei singoli amminoacidi in base all'altezza e al colore della lettera corrispondente.