

Ontologías y su importancia en Biología

Términos GO y Rutas KEGG

¿Qué es una Ontología?

Es una representación formal del conocimiento en la que los conceptos se describen por:

- Su significado
- Las relaciones que guardan entre ellos

A cada concepto o término se le asigna un identificador único (accession):

Accession: GO:0003700

Name: DNA-binding transcription factor activity

Description: A protein or a member of a complex that interacts selectively and non-covalently with a specific DNA sequence (sometimes referred to as a motif) within the regulatory region of a gene to modulate transcription. Regulatory regions include promoters (proximal and distal) and enhancers. Genes are transcriptional units, and include bacterial operons.

Ontology: molecular_function

¿Por qué son necesarias las Ontologías?

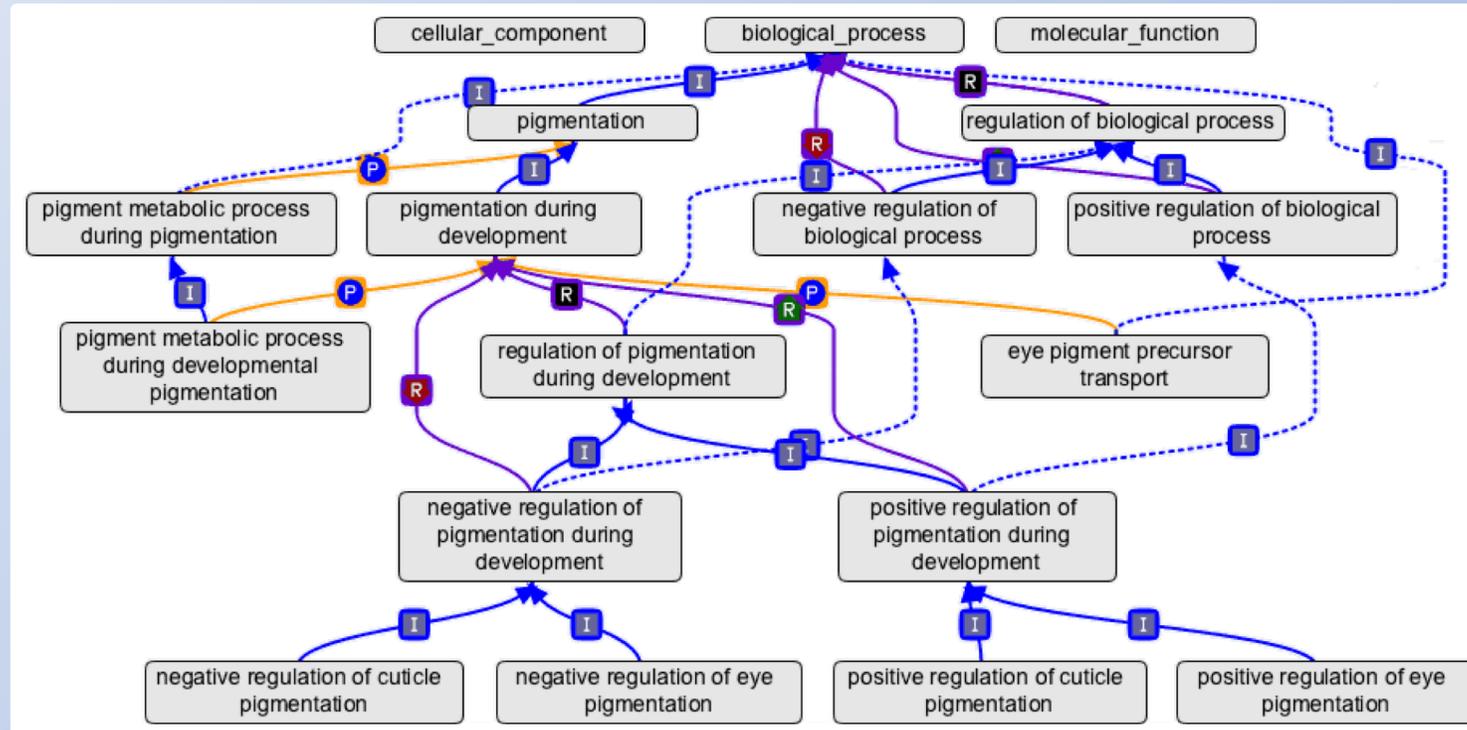
El conocimiento biológico es inherentemente complejo:

- Se puede dividir en categorías en función del objeto de estudio (procesos biológicos, funciones moleculares, componentes celulares, rutas metabólicas, rutas de transducción de señales, rutas reguladoras, etc.), pero todas estas categorías están entrelazadas.
- Los términos se relacionan jerárquicamente:
 - Un término más genérico (“término padre”) puede tener asociados uno o más términos más específicos (“términos hijo”).
 - Un término más específico (“término hijo”) puede tener asociados uno o más términos menos específicos (“términos padre”).
 - Los diferentes tipos de relaciones entre términos en la *Gene Ontology* están disponibles aquí: <http://www.geneontology.org/page/ontology-relations>

¿Por qué son necesarias las Ontologías?

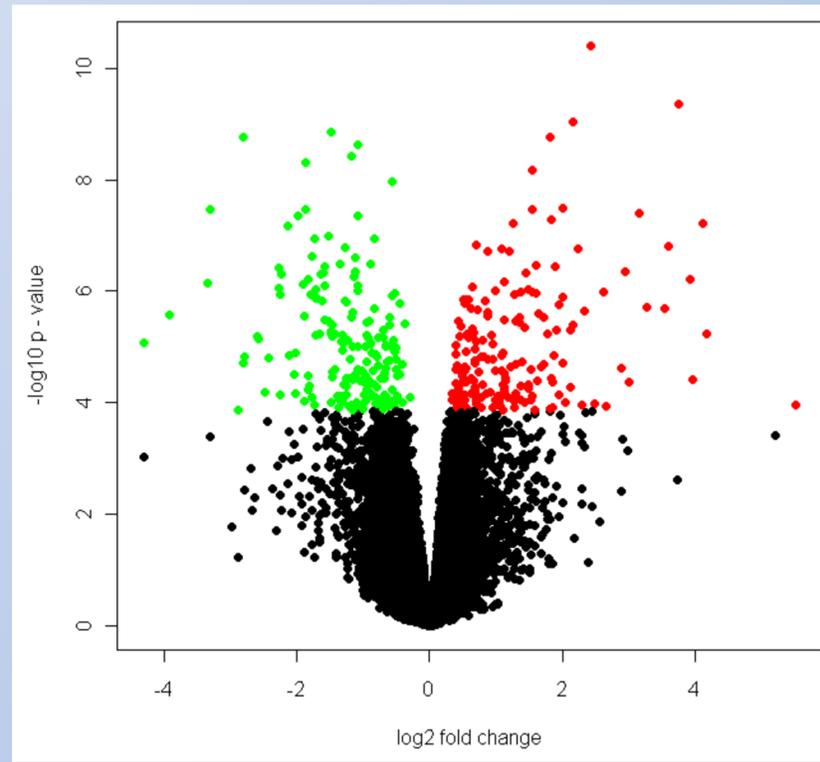
El conocimiento biológico es inherentemente complejo:

- Se puede dividir en categorías en función del objeto de estudio (procesos biológicos, funciones moleculares, componentes celulares, rutas metabólicas, rutas de transducción de señales, rutas reguladoras, etc.), pero todas estas categorías están entrelazadas.
- Los términos se relacionan jerárquicamente:



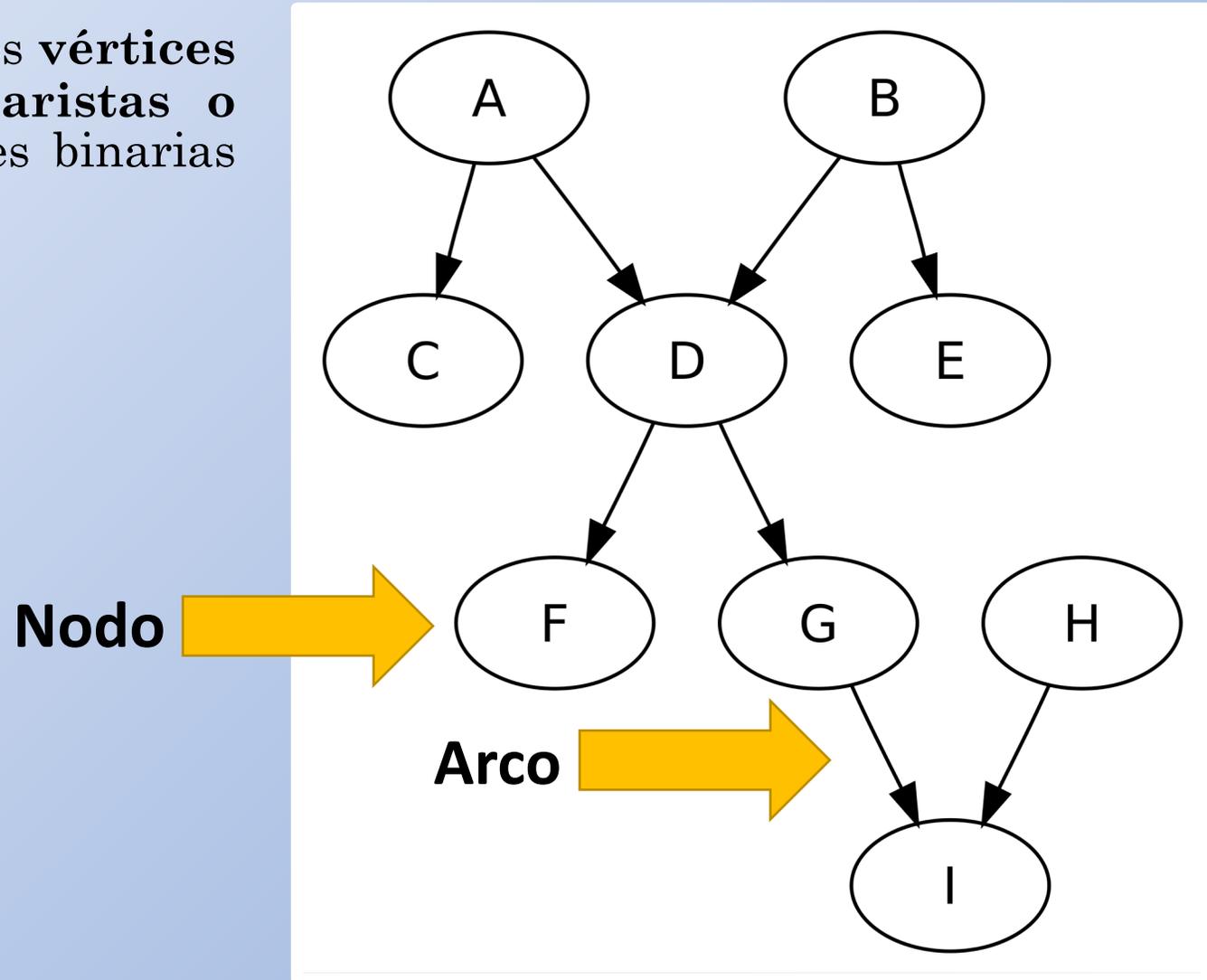
¿Por qué son necesarias las Ontologías?

- Las Ontologías son especialmente útiles para identificar qué tienen en común un conjunto de genes identificados como:
 - Genes diferencialmente expresados entre dos o más condiciones.
 - Genes diferencialmente afectados por marcas epigenéticas (ej. estado de la cromatina y metilación del promotor/potenciador).
 - Genes o proteínas implicados en procesos patológicos.



Grafo acíclico dirigido (DAG)

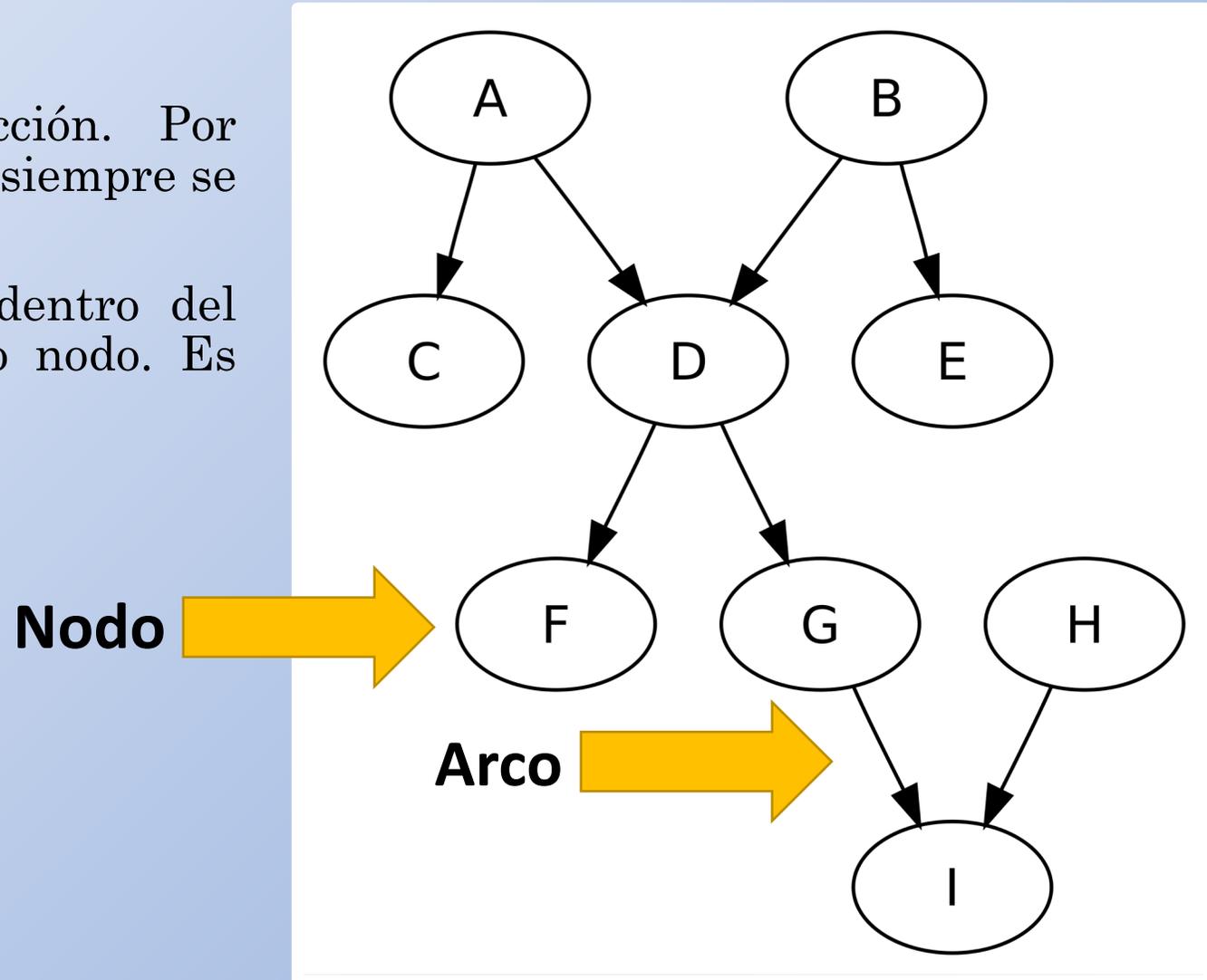
Un **grafo** es un conjunto de objetos llamados **vértices** o **nodos** unidos por enlaces llamados **aristas** o **arcos**, que permiten representar relaciones binarias entre elementos de un conjunto.



Grafo acíclico dirigido (DAG)

Un **DAG** es un tipo de grafo:

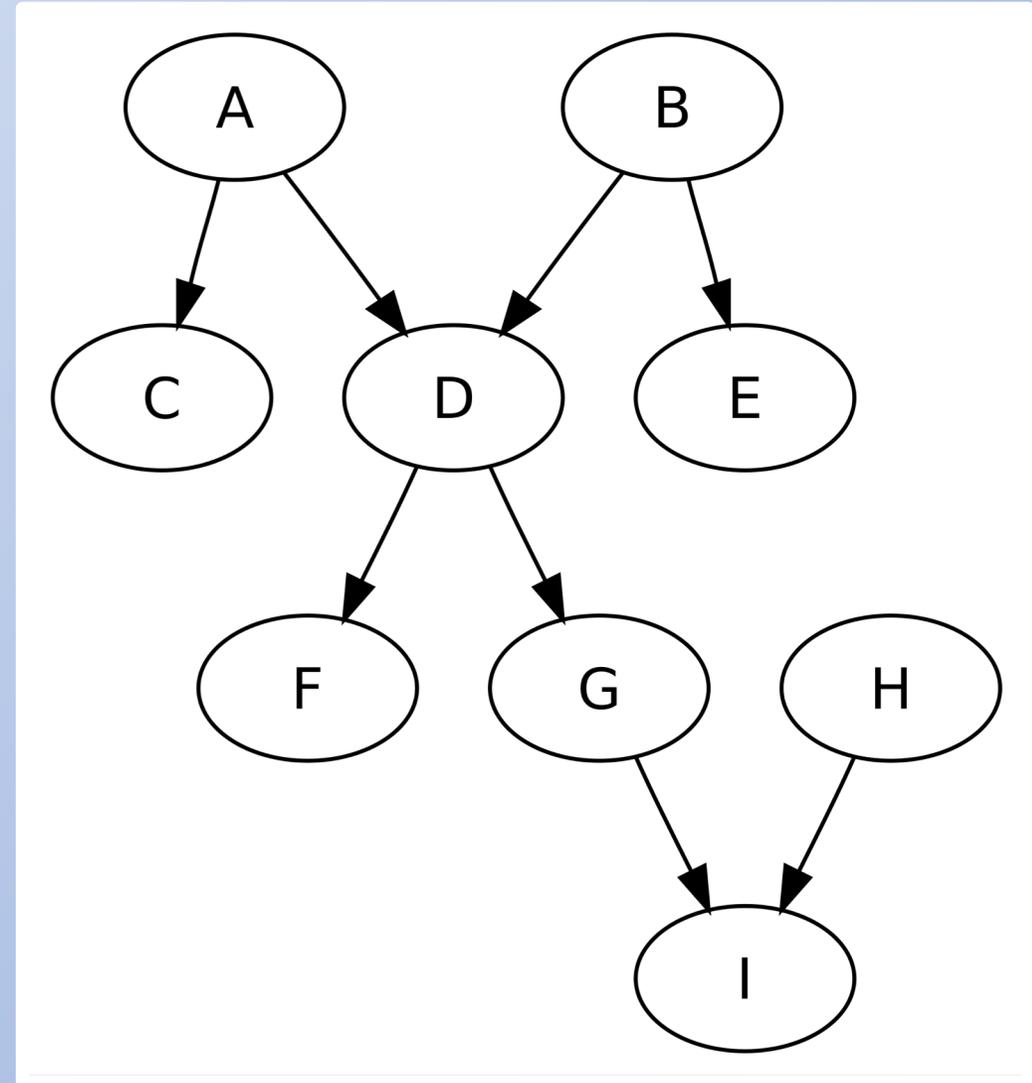
- **Dirigido:** los arcos indican la dirección. Por ejemplo, $G \rightarrow I$ (desde G hacia I). El DAG siempre se recorre siguiendo la dirección de los arcos.
- **Acíclico:** no existe un camino válido dentro del DAG que empiece y acabe en el mismo nodo. Es decir, el DAG no presenta ciclos.



Grafo acíclico dirigido (DAG)

En DAG de la derecha:

- El nodo A tiene dos nodos “hijo”: C y D
- El nodo B tiene dos nodos ”hijo”: D y E
- El nodo D tiene dos nodos “padre”: A y B
- El nodo D tiene dos nodos “hijo”: F y G
- El nodo G tiene un nodo “hijo”: I
- El nodo H tiene un nodo “hijo”: I
- El nodo I tiene dos nodos “padre”: G y H



Gene Ontology (GO)

El proyecto **Gene Ontology** es una iniciativa bioinformática con el objetivo de estandarizar la representación de los genes y los atributos de sus productos génicos de todas las especies:

- El proyecto proporciona un vocabulario controlado de términos y anotaciones de productos génicos.
- El sistema de términos está unificado y es consistente, de manera que cada término aparece solo una vez y se utiliza un diccionario de palabras permitidas, independiente de la especie.
- Los términos están relacionados unos con otros de manera jerárquica, desde términos más generales hacia términos más específicos.

Ontologías de GO

Proceso biológico:

Eventos celulares a los que contribuye el producto génico.

Función molecular:

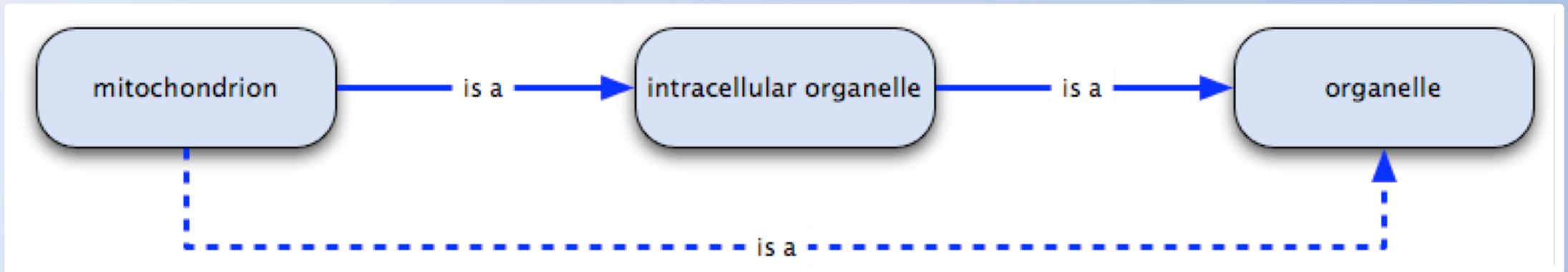
Descripción bioquímica del producto génico.

Componente celular:

Localización o complejos de los que forma parte el gene o su producto génico.

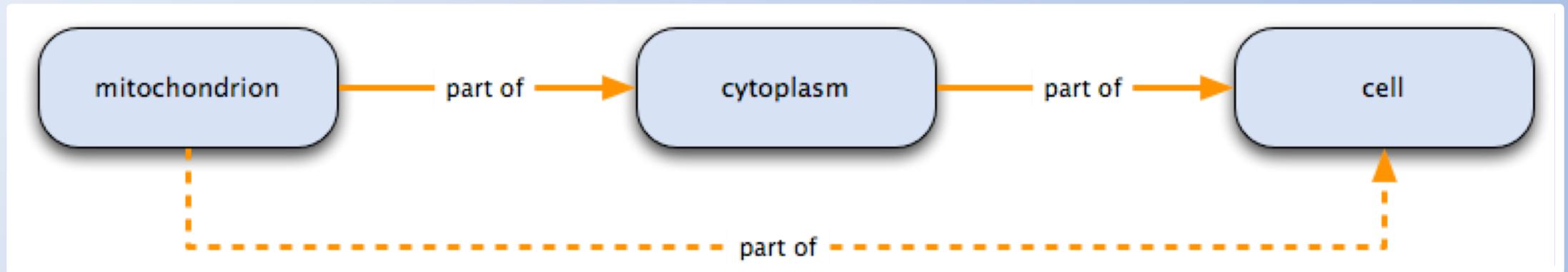
Tipos de arcos en GO

is a



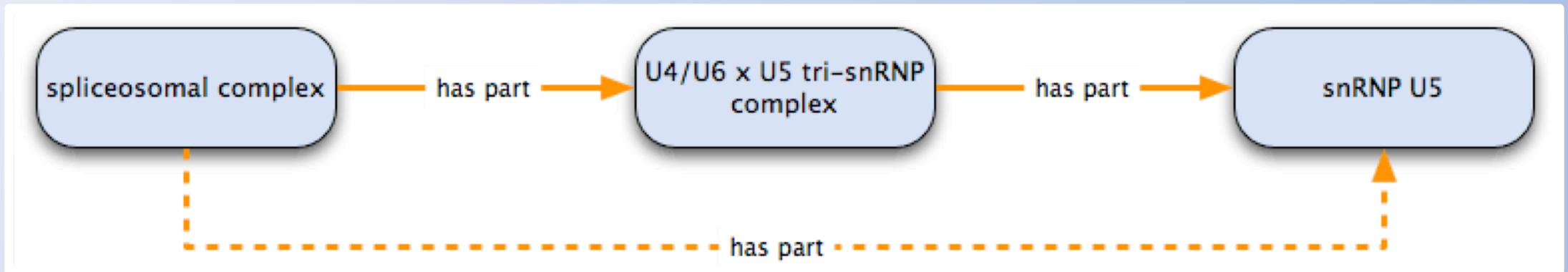
Tipos de arcos en GO

part of



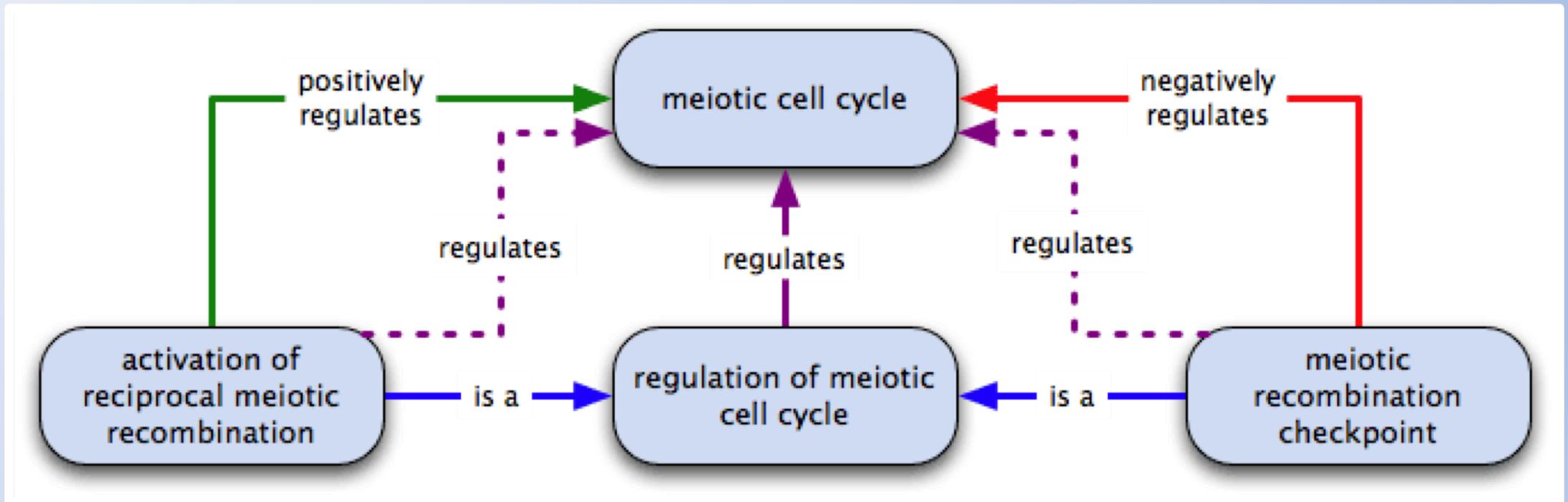
Tipos de arcos en GO

has part



Tipos de arcos en GO

regulates



Ejemplo de DAG en GO



Kyoto Encyclopedia of Genes and Genomes (KEGG)

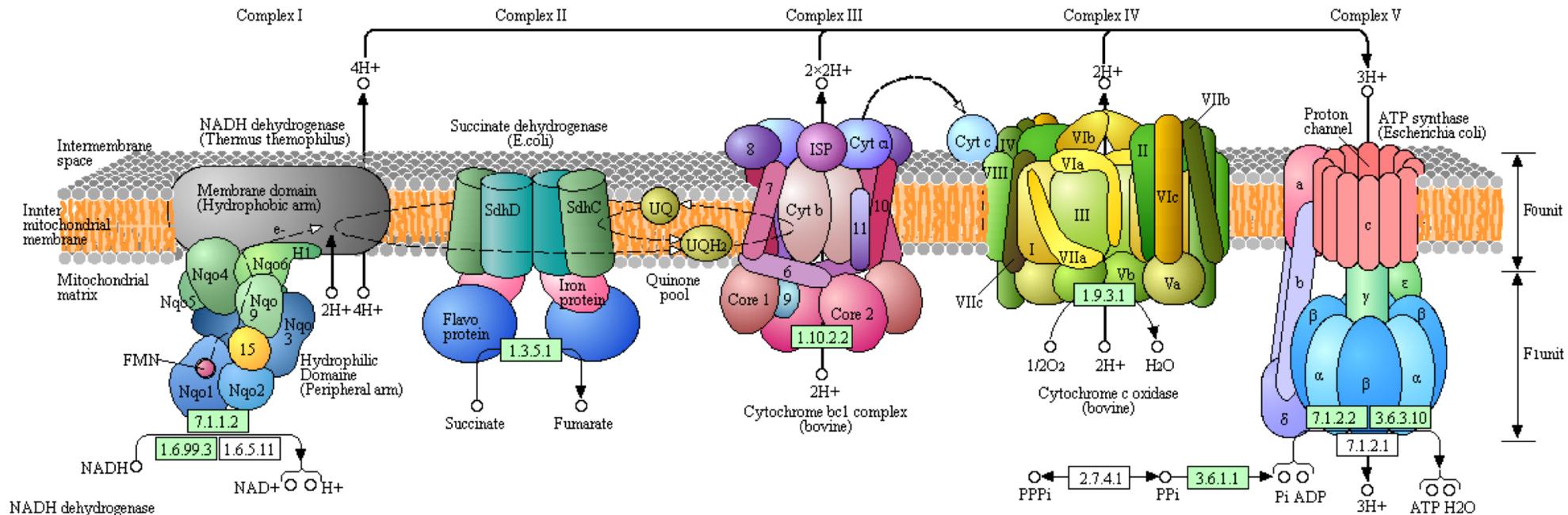
KEGG es un conjunto de base de datos diseñadas para facilitar la comprensión de sistemas biológicos (células, organismos o ecosistemas) a nivel molecular, reuniendo información procedentes de datos de secuenciación y otras tecnologías experimentales de alto rendimiento.

Su base de datos más conocida es **KEGG PATHWAY**: una colección de mapas de rutas dibujados manualmente y que representan nuestro conocimiento sobre interacciones moleculares, reacciones y redes de relación de:

1. Metabolismo
2. Procesado de la información genética
3. Procesado de la información ambiental
4. Procesos celulares
5. Sistemas orgánicos
6. Enfermedades humanas
7. Desarrollo de fármacos

OXIDATIVE PHOSPHORYLATION

Ejemplo de ruta en KEGG



NADH dehydrogenase

E	ND1	ND2	ND3	ND4	ND4L	ND5	ND6										
E	Ndufs1	Ndufs2	Ndufs3	Ndufs4	Ndufs5	Ndufs6	Ndufs7	Ndufs8	Ndufv1	Ndufv2	Ndufv3						
B/A	NuoA	NuoB	NuoC	NuoD	NuoE	NuoF	NuoG	NuoH	NuoI	NuoJ	NuoK	NuoL	NuoM	NuoN			
B/A	NdhC	NdhK	NdhJ	NdhH	NdhA	NdhI	NdhG	NdhE	NdhF	NdhD	NdhB	NdhL	NdhM	NdhN	HoxE	HoxF	HoxU
E	Ndufa1	Ndufa2	Ndufa3	Ndufa4	Ndufa5	Ndufa6	Ndufa7	Ndufa8	Ndufa9	Ndufa10	Ndufab1	Ndufa11	Ndufa12	Ndufa13			
E	Ndufb1	Ndufb2	Ndufb3	Ndufb4	Ndufb5	Ndufb6	Ndufb7	Ndufb8	Ndufb9	Ndufb10	Ndufb11	Ndufc1	Ndufc2				

Succinate dehydrogenase / Fumarate reductase

E	SDHC	SDHD	SDHA	SDHB		
B/A	SdhC	SdhD	SdhA	SdhB		
			FrdA	FrdB	FrdC	FrdD

Cytochrome c reductase

E/B/A	ISP	Cyt b	Cyt 1				
E	COR1	QCR2	QCR6	QCR7	QCR8	QCR9	QCR10

Cytochrome c oxidase

E	COX10	COX3	COX1	COX2	COX4	COX5A	COX5B	COX6A	COX6B	COX6C	COX7A	COX7B	COX7C	COX8	E/B/A	COX11	COX15	COX17
B/A	CyoE	CyoD	CyoC	CyoB	CyoA													
		CoxD	CoxC	CoxA	CoxB													
		QoxD	QoxC	QoxB	QoxA													

Cytochrome c oxidase, cbb3-type

B	I	II	IV	III
---	---	----	----	-----

Cytochrome bd complex

B/A	CydA	CydB	CydX
-----	------	------	------

F-type ATPase (Bacteria)

alpha	beta	gamma	delta	epsilon
a	b	c		

F-type ATPase (Eukaryotes)

alpha	beta	gamma	delta	epsilon	
OSCP	a	b	c	d	e
f	g	f6/h	j	k	8

V/A-type ATPase (Bacteria, Archaeas)

A	B	C	D	E	F	G/H
I	K					

V-type ATPase (Eukaryotes)

A	B	C	D	E	F	G	H
a	c	d	e	S1			

Términos GO vs. rutas KEGG

GO

- Visión panorámica.
- Se suele trabajar con un gran número de genes.
- Gran número de genes anotados.
- Cada gen es considerado por separado.

KEGG PATHWAY

- Visión en detalle.
- Se suele focalizar en un grupo concreto de genes.
- Solo algunos genes están anotados.
- Focaliza en la interacción entre genes (o entre sus productos génicos).