



# Banco de dados de vias metabólicas de biossíntese de terpenos em fungos e plantas

Waldeyr Mendes Cordeiro Silva

**Orientador:**

*Marcelo Macedo Brígido*

Agosto 2016

# Conceitos Básicos

## Metabolismo

- **Metabolismo:** conjunto de todas as reações que ocorrem em um organismo
- **Via metabólica:** cadeia de reações que podem ser catalisadas por **enzimas**, as quais frequentemente requerem co-fatores
- **Enzimas** originam-se do genoma (genes)
- **Metabolismo primário:** **essencial** para processos vitais, tais como crescimento e divisão celular, respiração e reprodução
- **Metabolismo secundário:** metabolismo de compostos **não essenciais** ao organismo

# Conceitos Básicos

## Metabólitos

Metabólitos são o substrato/produto das reações catalizadas por enzimas que ocorrem na célula e alguns critérios são necessários para definir um composto como metabólito, entre eles:

- Metabólitos são compostos encontrados nas células
- O produto de uma reação pode ser substrato para outra
- Metabólitos tem uma existência finita
- Metabólitos devem ter um papel biológico na célula
- Metabólitos podem ser reguladores do metabolismo

# Conceitos Básicos

## Tipos de metabólitos secundários

- **Policetídeos (PK)**
- **Peptídeos não-ribossomais (NRPS)**
- **Alcalóides**
- **Terpenóides**

# Conceitos Básicos

## Biossíntese de terpenóides

A biossíntese de terpenóides pode ser dividida em quatro estágios principais

1. formação de *isopentenyl diphosphate (IPP)*
2. condensação da unidade básica de isoprene  $C_5$  para gerar prenil-difosfatos
3. prenil-difosfatos ( $C_{10}$ ,  $C_{15}$ ,  $C_{20}$ ) submetidos a ciclizações e/ou rearranjos produzem esqueletos de carbono de cada classe de terpenos
4. formação dos terpenos a partir de oxidações e outras transformações dos esqueletos de carbono

# Conceitos Básicos

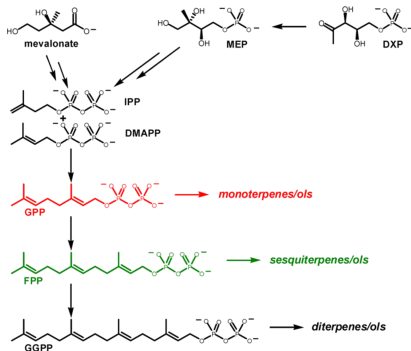


Fig. : Backbone da biossíntese de terpenóides<sup>a</sup>

<sup>a</sup>Wedler, H., Pemberton, R., & Tantillo, D. (2015). Carbocations and the Complex Flavor and Bouquet of Wine: Mechanistic Aspects of Terpene Biosynthesis in Wine Grapes. *Molecules*, 20(6), 10781–10792.

# Conceitos básicos

## Papel ecológico

- agentes anti-microbianos
- agentes anti-herbívoros
- agentes inseticidas
- sinalização em relacionamentos ecológicos
  - atração de polinizadores
  - atração de dispersadores de sementes
  - relacionamentos planta-planta, planta-animal, planta-micróbios

## Conceitos básicos

A word cloud on a white background with a soft drop shadow. The words are arranged in a roughly circular pattern, all tilted at an angle. The word 'stress' is the largest and most prominent, located at the bottom center. Other words include 'sementes' (large, above stress), 'inseticida' (above sementes), 'anti-herbívoro' (above inseticida), 'comunicação' (top left), 'antivírus' (top right), 'fungicida' (left side), 'sinalização' (bottom right), and 'bactericida' (bottom right, below sinalização).

comunicação  
antivírus  
inseticida  
anti-herbívoro  
sementes  
fungicida  
**stress**  
sinalização  
bactericida

Fig. : Papel ecológico dos terpenos, palavras-chave em revisão de literatura 2010-2016.



# Conceitos básicos

## Como representar e armazenar essas informações?

	Banco de dados factográfico	Arquivo estruturado	Grafo	Banco de dados relacional	Rede de Petri	Programação Lógica	Não identificado	GEM [Arakawa2006]
Selkov1989	X							
Baher1992					X			
Karp1994		X						
Reddy1993				X				
Gaasterland1995					X			
Kanehisa2000						X		
Hucka2001		X						
Goesmann2002			X					
Pharkya2002						X		
Boyer2003						X		
Forster2003		X						
Pinney2003				X				
Herrgard2008								X
Grossetete2010			X					
Dreyfuss2013		X	X					

## Grafos

Um grafo  $G$  é um conjunto não vazio de vértices  $V(G)$  e um conjunto finito de arestas  $E(G)$  disjunto de  $V(G)$  e uma função incidental  $\psi_G$  que associa a cada aresta de  $G$  um par não ordenado e não necessariamente distinto de vértices de  $G^a$ .

$$V = \{V_1, V_2, \dots, V_n\}$$

onde  $V$  é o conjunto de vértices  $V_1, V_2, \dots, V_n$

$$E = \{e_1, e_2, \dots, e_m\}$$

onde  $E$  é o conjunto de arestas  $e_1, e_2, \dots, e_m$

$$\psi_G(e_k) = \{V_i, V_j\}$$

onde  $V_i, V_j \in V$  e  $e_k$  é a aresta que liga  $V_i$  a  $V_j$

---

<sup>a</sup>Bondy, John Adrian, and Uppaluri Siva Ramachandra Murty. Graph theory with applications. Vol. 290. London: Macmillan, 1976.

# Conceitos básicos

## NoSQL

Os bancos de dados NoSQL surgiram como uma alternativa para grandes volumes de dados devido às suas características, tais como escalabilidade e flexibilidade.

O termo NoSQL (Not Only SQL) se refere a estes novos bancos de dados que podem ser classificados como *Key-Value*, *Big Table*, Documentos e Grafos.

## Problema

Integrar informações sobre biossíntese de terpenos a partir dos bancos de dados KEGG<sup>a</sup>, BioCyc<sup>b</sup>, ChEBI<sup>c</sup> e revisão de literatura em um banco de dados em grafos e disponibilizá-lo para a comunidade científica.

---

<sup>a</sup><http://www.kegg.jp>

<sup>b</sup><http://biocyc.org>

<sup>c</sup><https://www.ebi.ac.uk/chebi>

## Hipótese

Um novo banco de dados especializado em metabolismo de terpenóides facilitará a pesquisa nesta área ao mesmo tempo que permitirá usar os dados bioquímicos para predições de novas vias para síntese de um dado terpeno.

## Objetivo Principal

Construir e prover acesso a um novo banco de dados especializado em metabolismo de terpenos em plantas e fungos que preserve características individuais da biossíntese destes compostos nesses organismos.

## Objetivos específicos

- Criar um modelo de banco de dados capaz de representar vias metabólicas em plantas e fungos com características como compartimentação celular, transporte e armazenamento dos metabólitos;
- Obter e reorganizar informações sobre metabolismo de terpenos nos bancos de dados KEGG e BioCyC programaticamente via *web service*;
- Obter informações sobre metabolismo de terpenos na literatura via revisão bibliográfica;
- Organizar toda a informação obtida em um banco de dados em grafos de acordo com o modelo criado representá-las;
- Disponibilizar a consulta a essas informações via *web service*.

# Trabalhos relacionados



Database, 2014, 1–12  
doi: 10.1093/database/bat100  
Original article



BIOENGINEERING AND BIOTECHNOLOGY

METHODS ARTICLE  
published: 12 October 2014  
doi: 10.1093/database/bat100



## Original article

### EssOilDB: a database of essential oils reflecting terpene composition and variability in the plant kingdom

Sangita Kumari, Sachin Pundhir, Piyush Priya, Ganga Jeena, Ankita Punetha, Konika Chawla, Zohra Firdos Jafaree, Subhasish Mondal and Gitanjali Yadav\*

International Journal of Genomics  
Volume 2015 (2015), Article ID 502795, 7 pages  
<http://dx.doi.org/10.1155/2015/502795>

## Research Article

### Evaluating the Cassandra NoSQL Database Approach for Genomic Data Persistence

Rodrigo Aniceto,<sup>1</sup> Rene Xavier,<sup>1</sup> Valéria Guimarães,<sup>1</sup> Fernanda Hondo,<sup>1</sup> Maristela Holanda,<sup>1</sup> Maria Emilia Walter,<sup>1</sup> and Sérgio Lifschitz<sup>2</sup>

## RESEARCH ARTICLE

### miTALOS v2: Analyzing Tissue Specific microRNA Function

Martin Preusse<sup>1,2</sup>, Fabian J. Thels<sup>1,3</sup>, Nikola S. Mueller<sup>1\*</sup>

<sup>1</sup> Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, <sup>2</sup> Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, <sup>3</sup> Institute for Mathematical Sciences, Technische Universität München, Munich, Germany

\* [nikola.mueller@helmholtz-muenchen.de](mailto:nikola.mueller@helmholtz-muenchen.de)

### Comprehensive reconstruction and visualization of non-coding regulatory networks in human

Vincenzo Bonnici<sup>1</sup>, Francesco Russo<sup>2,3</sup>, Nicola Bombieri<sup>1</sup>, Alfredo Pulvirenti<sup>4,\*</sup> and Rosalba Giugno<sup>4,\*</sup>

<sup>1</sup> Department of Computer Science, University of Bari, Bari, Italy  
<sup>2</sup> Laboratory of Integrative Systems Medicine (LISM), Institute of Informatics and Telematics (IIT) and Institute of Clinical Physiology (IRC), National Research Council (CNR), Pisa, Italy  
<sup>3</sup> Department of Computer Science, University of Pisa, Pisa, Italy  
<sup>4</sup> Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy

Simões et al. BMC Bioinformatics 2015, 16(Suppl 1):S59  
<http://www.biomedcentral.com/1471-2105/16/S1/S59>



## RESEARCH

## Open Access

### NERI: network-medicine based integrative approach for disease gene prioritization by relative importance

Sérgio N Simões<sup>1,2</sup>, David C Martins Jr<sup>3</sup>, Carlos Ali Pereira<sup>1</sup>, Ronaldo F Hashimoto<sup>1</sup>, Helena Brentani<sup>4,5,6</sup>

From Brazilian Symposium on Bioinformatics 2014

Bioinformatics Advance Access published October 21, 2015

## BIOINFORMATICS

## EDITORIAL

2015, pages 1–2  
doi:10.1093/bioinformatics/bt0549

## Databases and ontologies

Advance Access publication October 17, 2015

### Are graph databases ready for bioinformatics?

Christian Theil Have<sup>1</sup> and Lars Juhl Jensen<sup>2,\*</sup>

<sup>1</sup>Department of Metabolic Genetics, Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, 2100 Copenhagen Ø, Denmark and <sup>2</sup>Department of Disease Systems Biology, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, 2100 Copenhagen N, Denmark

Associate Editor: Jonathan Wren

### CGDM: Collaborative Genomic Data Model for Molecular Profiling Data Using NoSQL



Shicai Wang<sup>1</sup>, Mihaela A. Mares<sup>1</sup> and Yi-ke Guo<sup>1,2,\*</sup>

+ Author Affiliations

\* To whom correspondence should be addressed. Yi-ke Guo, E-mail: [y.guo@imperial.ac.uk](mailto:y.guo@imperial.ac.uk)

Received April 18, 2016.  
Revision received July 21, 2016.  
Accepted August 9, 2016.

## Método

1. Conceber e modelar o banco de dados em grafo
2. Implementar o modelo de dados em um banco de dados em grafos chamado (2Path) usando o (OrientDB<sup>a</sup>)
3. Construir uma aplicação para recuperar as informações dos bancos de dados via *web service* e popular automaticamente o 2Path
4. Construir uma interface para popular o 2Path com dados da literatura
5. Construir uma API (*Application Program Interface*) para acesso ao 2Path via *web service*

---

<sup>a</sup><http://www.orientdb.com>



## Diferenças entre bancos

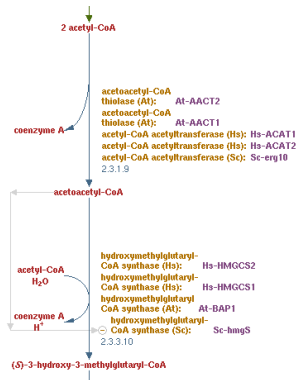
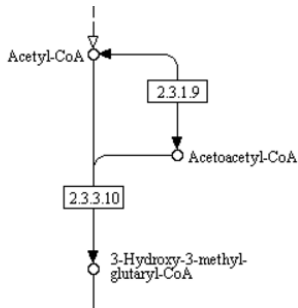


Fig. : Diferença entre KEGG e Biocyc para a biossíntese de Hydroxymethylglutaryl-CoA.

## Compartimentação celular

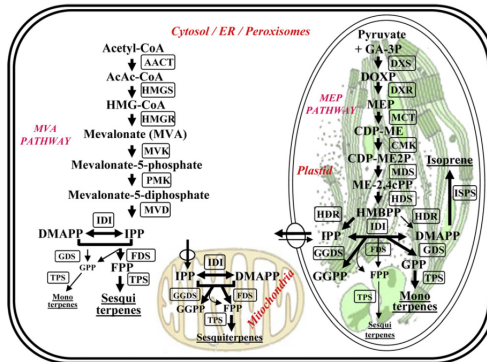


Fig. : Transporte inter-celular da molécula de IPP<sup>a</sup>.

<sup>a</sup> Nagegowda, D. A. (2010). Plant volatile terpenoid metabolism: Biosynthetic genes, transcriptional regulation and subcellular compartmentation. *FEBS Letters*, 584(14), 2965–2973

# Resultados iniciais

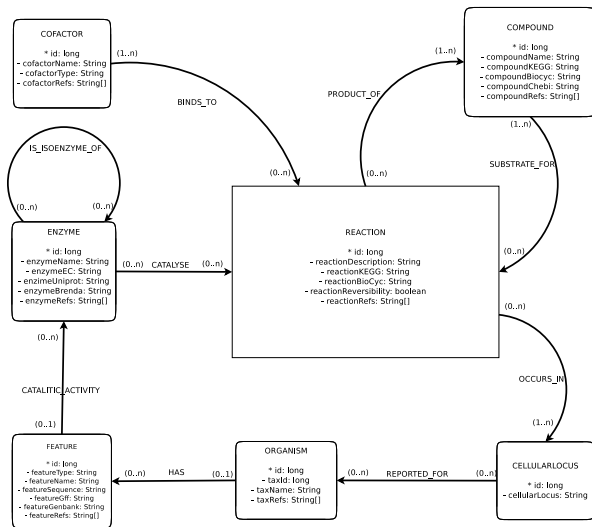


Fig. : Modelo de dados proposto para o 2Path usando a notação GRADED (Graph Description Diagram for Graph Databases)

## Agenda - atividades e cronograma

1. Adaptar o software para o OrientDB
2. Popular o 2Path com os dados do KEGG/BioCyc/CHEBI
3. Desenvolver a interface para inserir dados da literatura
4. Popular o 2Path com literatura para a 1ª fase da biossíntese
5. Popular o 2Path com literatura para a 2ª e 3ª fase da biossíntese
6. Estudar conceitos de gramática de grafos
7. Popular o 2Path literatura para a 4ª fase da biossíntese
8. Implementar predição de vias usando gramática de grafos

	2016-2	2017-1	2017-2	2018-1	2018-2
1	X				
2	X				
3	X				
4	X				
5		X			
6		X	X		
7		X	X		
8			X	X	



# Banco de dados de vias metabólicas de biossíntese de terpenos em fungos e plantas

Waldeyr Mendes Cordeiro Silva

**Orientador:**

*Marcelo Macedo Brígido*

Agosto 2016