

Estudo da Plasticidade Evolutiva Através da Distribuição de Grupos Ortólogos

Luís H. T. de Souza, Rodrigo J. S. Dalmolin, Mauro A.A. Castro, José L. Rybarczyk-Filho, Rita M.C. de Almeida, José CF Moreira

Objetivo

Propor uma metodologia capaz de avaliar plasticidade evolutiva, através de um índice, contemplando unidades evolutivas que extrapolassem nível gênico. O Índice de Plasticidade Evolutiva baseia-se então na distribuição de grupos de ortólogos ao longo dos eucariotos.

Fundamentação

Supõe-se que genes de grande importância aos processos metabólicos sejam menos sujeitos a modificações, dificultando portanto, o estabelecimento de novas cópias.

Supõe-se também, que dada sua importância, a distribuição de seus parálogos ao longo dos organismos seja bastante homogênea.

Abundância

$$\frac{\text{Número de Proteínas de "a"}}{\text{Número de Organismos contendo "a"}}$$

Diversidade

$$-\frac{1}{\ln(\text{Número total de organismos})} \sum_i p(i,a) \ln p(i,a)$$

O Índice de Plasticidade Evolutiva

$$\text{Índice de Plasticidade Evolutiva de "a"} = 1 - \frac{\text{Diversidade } [H_a]}{\sqrt{\text{Abundância } [D_a]}}$$

Fig. 4: Equação de determinação do Índice de Plasticidade Evolutiva, baseando-se nos índices de Diversidade e Abundância.

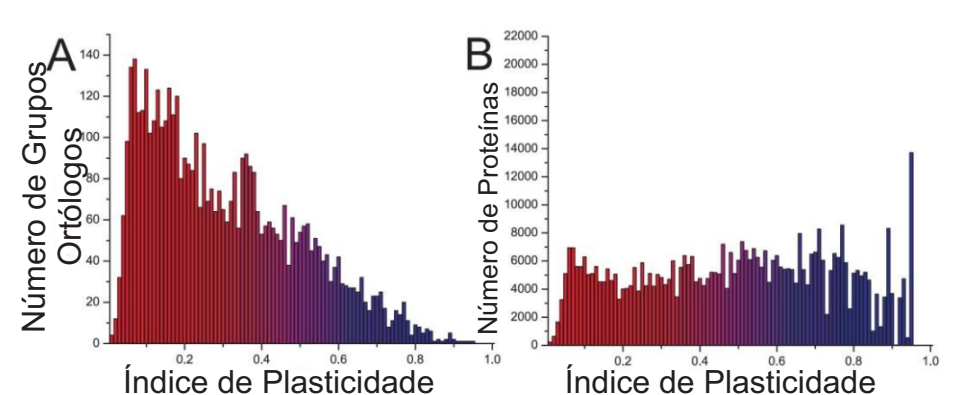


Fig. 5: Distribuição dos grupos ortólogos (A) e Proteínas (B) quanto ao valor de seu Índice de Plasticidade Evolutiva.

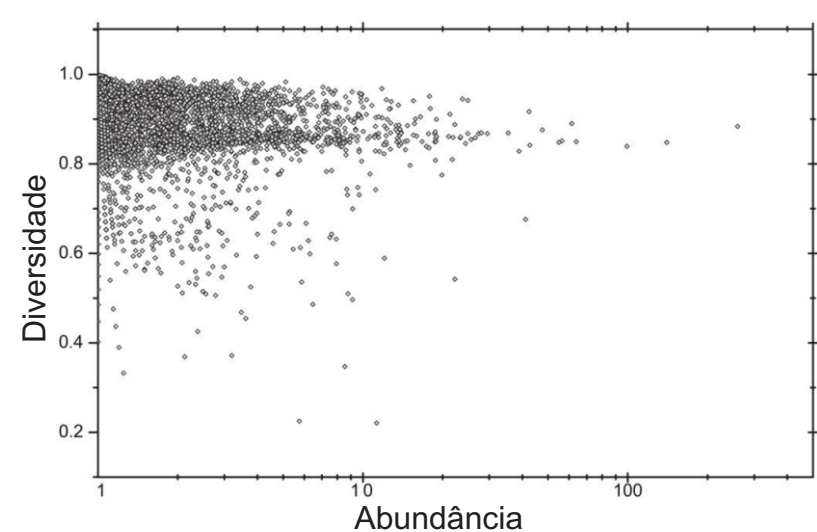


Fig. 3: Distribuição dos grupos ortólogos quanto a seus valores de abundância e diversidade.

"a" – leia-se "determinado agrupamento de grupos ortólogos"; // "p" – Probabilidade; // "i" – Item (gene/proteína) pertencente a "a";

Metodologia

Avaliou-se os índices distribuição de 4850 grupos ortólogos eucariotos (compreendendo 481.421 proteínas) ao longo de 55 organismos modelo. Os dados foram obtidos da base online do NCBI (www.ncbi.nlm.nih.gov/COG/).

Os índices de abundância e diversidade foram computados com o auxílio do software GenPlast*, que com estes compõe o Índice de Plasticidade Evolutiva.

Para validação do Índice, recorreu-se: a base de dados do Kegg (kegg.jp) de onde se obteve as informações dos genes que compunham as rotas metabólicas, cujo diagrama de interações, gerado no STRING (string.embl.de) então processado no ViaComplex* onde o diagrama (a rede) recebe uma camada de cor correspondente ao seu valor de plasticidade. Recorreu-se também ao Saccharomyces Genome Database (yeastgenome.org) e o Mouse Genome Informatics (informatics.jax.org), de onde se obteve as informações de fenótipo resultante do "nockout" de cada um dos genes. Analisou-se a plasticidade dos genes cuja deleção levava a inviabilidade do organismo, perante os demais (que não inviabilizavam o organismo e os totais.)

*Ambos softwares desenvolvidos "na casa".

Validação

O Índice de Plasticidade Evolutiva foi desafiado pelos 3 seguintes questionamentos:

- 1- Haveria **relação** entre o Índice de Plasticidade Evolutiva e a **distância evolutiva** entre os genes?
- 2- Genes **essenciais a sobrevivência** apresentariam um **baixo Índice de Plasticidade Evolutiva**?
- 3- Os **processos metabólicos** referenciados empiricamente como **conservados** apresentariam um **baixo Índice de Plasticidade Evolutiva** em seus genes?

1: Correlação

Analisou-se a correlação entre o Índice de Plasticidade Evolutiva e os valores de distância evolutiva computado para cada grupo ortólogo em função da distância evolutiva dos genes que pertencem ao grupo.

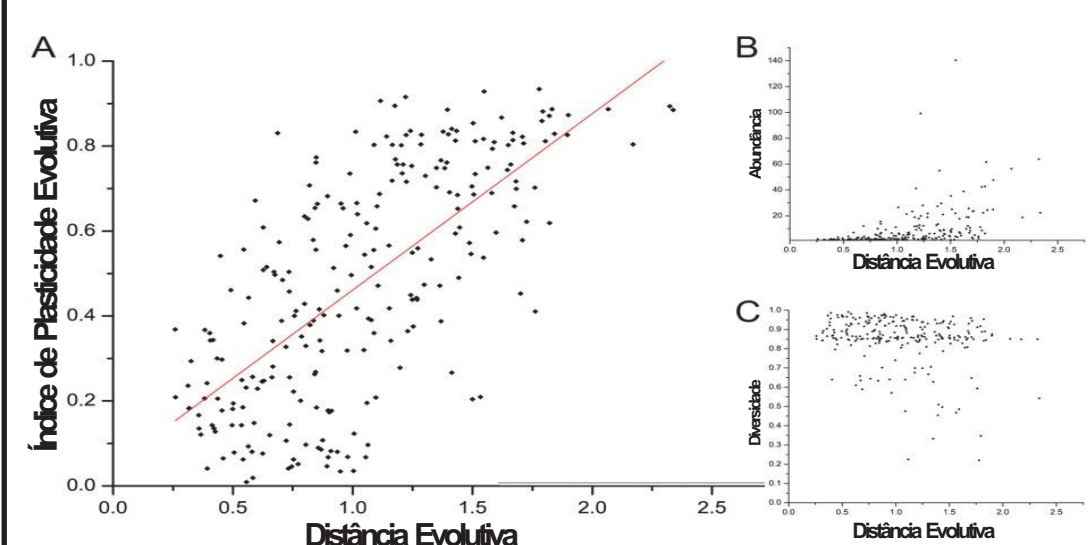


Fig 7: (A) Correlação ($P < 0,001$, $R^2 = 0,47$) entre o índice de Plasticidade Evolutiva e a Distância evolutiva de uma amostra de 5% dos grupos ortólogos do STRING; Os gráficos B e C mostram que não há correlação com os valores de Abundância e Diversidade em separado.

2: Essencialidade

Avaliou-se a plasticidade de genes que quando "nockoutados" em *S. cerevisiae* e *M. musculus*, levariam a inviabilidade do organismo:

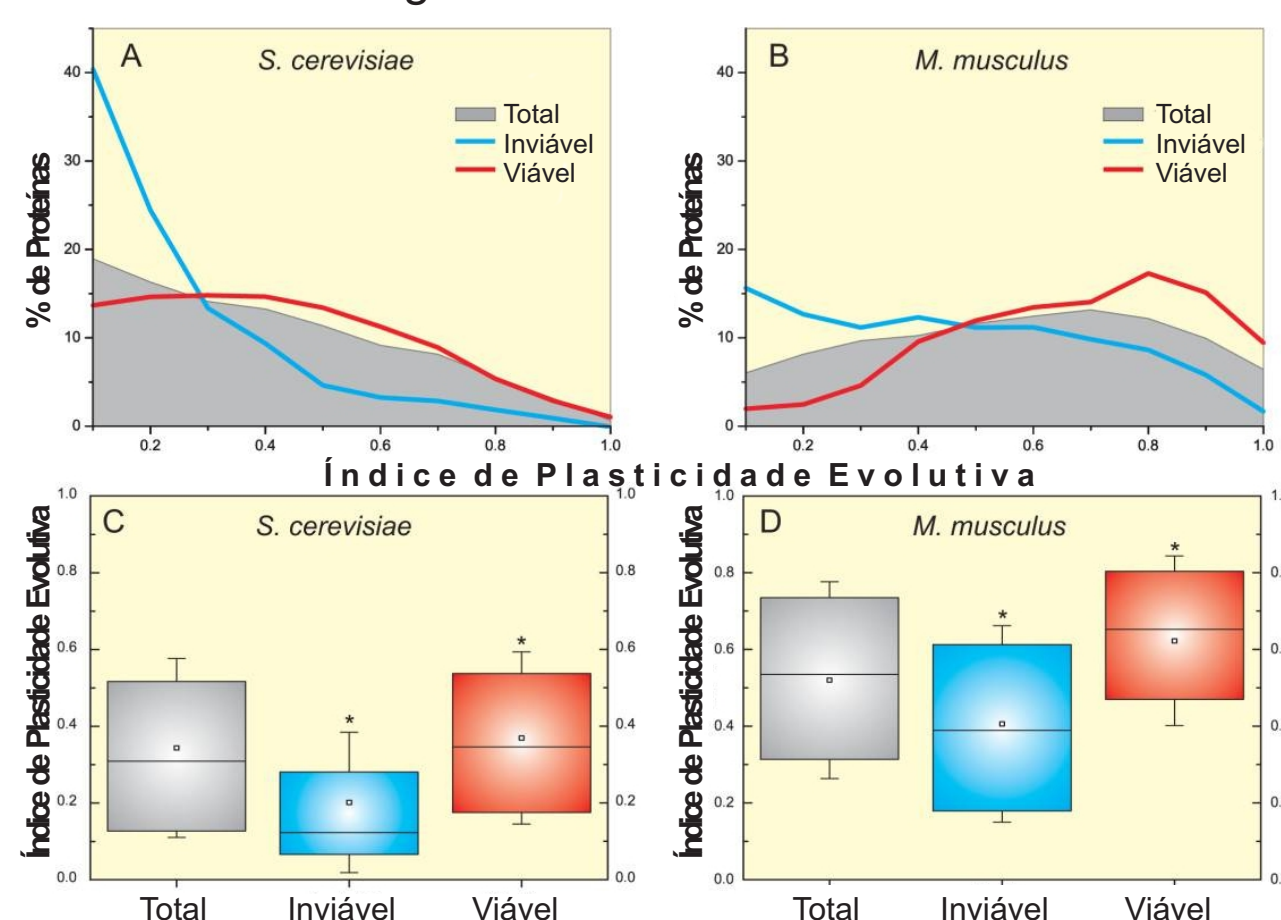


Fig 8: Distribuição de proteínas quanto aos seus valores de Plasticidade, relativo aos grupos de genes que levam a inviabilidade da levedura (A) ou Camundongo (B) quando nockoutados. "Chartbox" (C e D) - $P < 0,0001$ para ambos - analisando estatisticamente a distribuição dos genes associados a cada fenótipo, para *S. cerevisiae* (C) e *M. Musculus* (D).

3: Processos "Conservados"

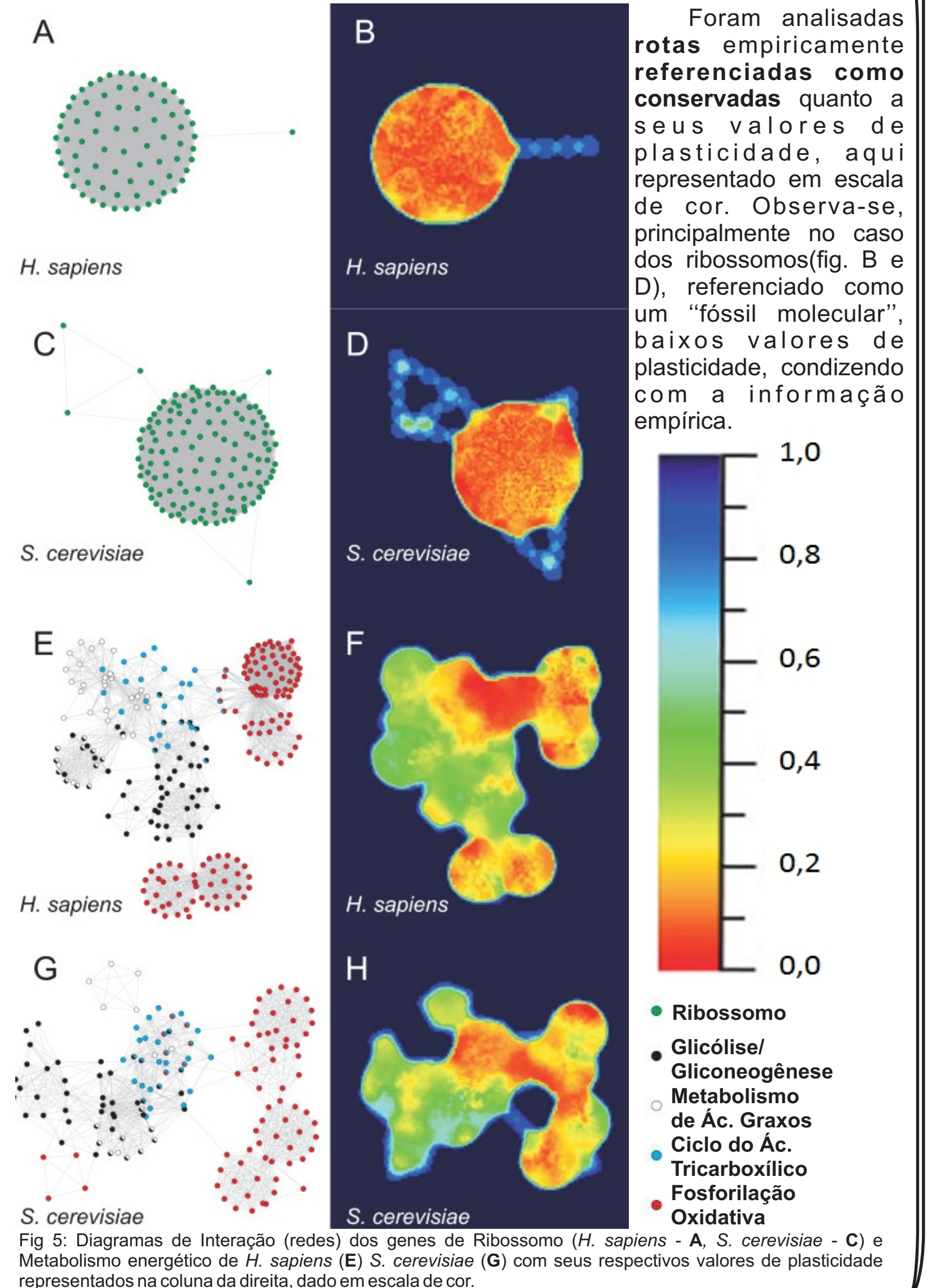


Fig 5: Diagramas de Interação (redes) dos genes de Ribossomo (*H. sapiens* - A, *S. cerevisiae* - C) e Metabolismo energético de *H. sapiens* (E) *S. cerevisiae* (G) com seus respectivos valores de plasticidade representados na coluna da direita, dado em escala de cor.

Conclusão

A análise da distribuição de grupos ortólogos é capaz de inferir, por meio de um valor numérico, o quão plástica ou conservada é uma determinada família gênica. O índice mostrou-se coerente perante medidas evolutivas inferidas a partir de sequência de aminoácidos, bem como em referenciar a baixa plasticidade de genes essenciais e/ou envolvidos em rotas metabólicas empiricamente referenciadas como 'conservadas'. O Índice de Plasticidade Evolutiva portanto parece ser eficiente em discriminar entre famílias gênicas mais ou menos estáveis ao longo da evolução.