

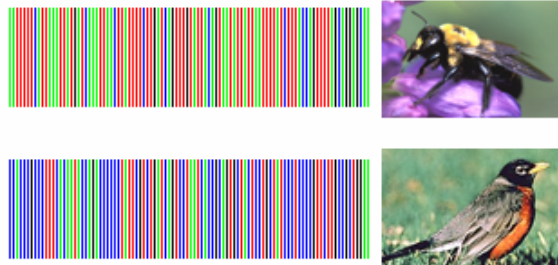
# O CÓDIGO DE BARRAS DA VIDA, ILUSTRADO

## Objetivos, Justificativa, Resultados

Mark Stoeckle    The Rockefeller University  
Paul E. Waggoner    Connecticut Agricultural Experiment Station  
Jesse H. Ausubel    Alfred P. Sloan Foundation

v1.3

24 de janeiro de 2005



**Tradução: Cristina Yumi Miyaki    Universidade de São Paulo**  
**18 may 2007**

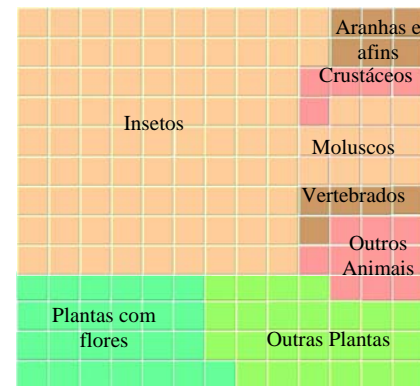
CÓDIGO DE BARRAS DA VIDA: Uma seqüência curta de DNA com localização uniforme nos genomas utilizada para identificar espécies.

A obtenção de códigos de barras é uma metodologia padronizada para identificar animais e plantas a partir de uma seqüência mínima de DNA.

1. Por que obter o código de barras de animais e de plantas?

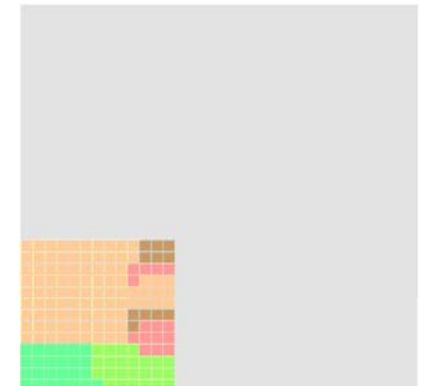
Pela integração dos avanços em eletrônica e em genética, a obtenção de código de barras auxiliará, de modo rápido e barato, identificar e obter informações sobre espécies descritas. Além disso, milhões de novas espécies ainda não descritas poderão ser identificadas. O código de barras fornecerá novos meios para apreciar e manejar a enorme e dinâmica biodiversidade na Terra.

Biodiversidade conhecida  
(excluindo microorganismos)  
Aproximadamente 1,7 milhões de espécies  
conhecidas de plantas e de animais



1 quadrado = 10.000 espécies

Biodiversidade estimada  
(excluindo microorganismos)  
10 milhões de espécies

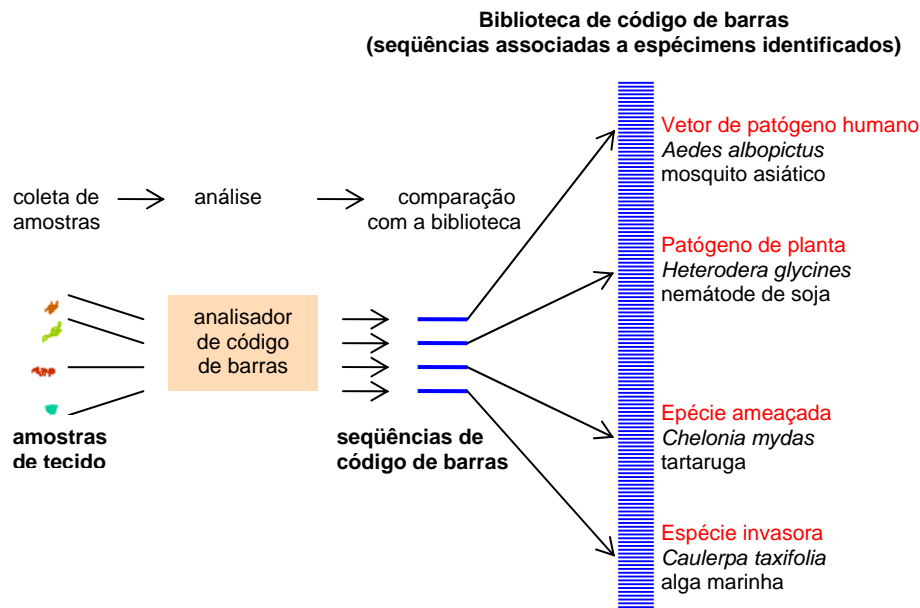


## 2. Quais são os benefícios da padronização?

Os técnicos têm desenvolvido várias maneiras de identificar espécies utilizando o DNA, geralmente adaptando a metodologia para responder uma pergunta específica para um grupo limitado de espécies. A metodologia do código de barras visa obter os benefícios da padronização. A padronização geralmente diminui custos e aumenta a confiabilidade, resultando em maior difusão e uso.

No caso do código de barras, a padronização deve auxiliar a acelerar a construção de uma biblioteca de referência consistente e abrangente de seqüências de DNA e o desenvolvimento de tecnologias econômicas para a identificação de espécies. O objetivo é permitir que qualquer pessoa, em qualquer lugar, a qualquer momento possa identificar rapidamente e com precisão a espécie de um determinado espécimen em qualquer condição de preservação.

Até o momento os resultados sugerem que um gene mitocondrial permitirá identificar a maioria das espécies de animais. No caso de espécies próximas de plantas, o nível de diferenciação encontrado em genes mitocondriais não permite a sua distinção. A metodologia para padronizar a identificação de plantas, usando uma ou possivelmente duas regiões genômicas, está sendo desenvolvida.



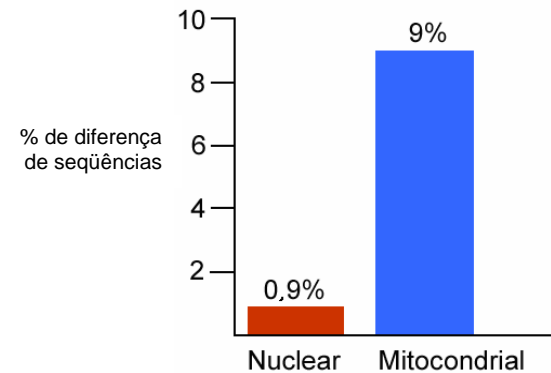
## 3. Por que usar o DNA mitocondrial como código de barras da vida?

As mitocôndrias, organelas que produzem energia nas células de plantas e de animais, possuem seu próprio genoma. Seqüências de DNA mitocondrial têm sido utilizadas há vinte anos em estudos de diferenciação de espécies próximas de animais. Quatro características tornam o genoma mitocondrial especialmente útil na identificação de espécies:

**Número de cópias.** Em geral uma célula possui duas cópias de seqüências nucleares e de 100 a 10000 cópias do genoma mitocondrial. O sucesso de obtenção de seqüências mitocondriais é geralmente muito maior do que o de seqüências nucleares, especialmente em amostras diminutas ou degradadas. Maior sucesso com amostras limitadas resulta em custos de processamento menor.

**Maior diferenciação entre espécies.** A diferença de seqüências entre espécies próximas são 5 a 10 vezes maiores em genes mitocondriais do que em genes nucleares. Assim, curtos segmentos de DNA mitocondrial podem distinguir espécies e, como são curtas, são mais baratas.

Diferença média observada no DNA nuclear e mitocondrial entre o homem e o chimpanzé



**Poucas diferenças dentro da espécie.** A variação intraespecífica do DNA mitocondrial é pequena na maioria das espécies animais. Isso pode ser consequência da rápida perda de polimorfismos ancestrais devido à herança materna ou a varredura seletiva após a origem de mutações vantajosas. Deixando de lado as causas, diferenças intraespecíficas pequenas e interespecíficas grandes resultam em limites genéticos entre a maioria das espécies, permitindo a sua identificação precisa baseado em um código de barras mitocondrial.

*Ausência de íntrons.* Os genes mitocondriais de animais raramente possuem íntrons (seqüências não codificadoras dispersas entre regiões codificadoras de um gene). Assim, é geralmente fácil obter a amplificação de DNA mitocondrial. Em contraste, a amplificação de regiões codificadoras de genes nucleares pode ser dificultada pela presença de íntrons, que podem ser longos.

#### 4. Quais são as maiores limitações encontradas até o momento no uso do código de barras?

*Grupos com baixa diversidade de seqüências.* Um exemplo foi encontrado em alguns corais e anêmonas do filo marinho Cnidária. A prevalência desses grupos não é conhecida, pois apenas algumas espécies foram analisadas e seqüências de DNA mitocondrial são distintas entre algumas espécies próximas pertencentes a esse grupo. Uma comparação entre seqüências mitocondriais de 2238 espécies de 11 filos animais revelou que 98% de pares de espécies próximas apresentam mais de 2% de diferenças em suas seqüências, o que seria suficiente para identificar corretamente a maioria das espécies.

*Resolução entre espécies com divergência recente.* A separação de organismos proximamente relacionados que adquiriram status de espécies recentemente é difícil de ser obtida usando qualquer método, inclusive morfologia. Em alguns casos, o código de barras mitocondrial pode chegar ao limite de identificação entre duas (ou mais) espécies próximas. A frequência de espécies com o mesmo código de barras é baixa dentro dos grupos analisados até o momento.

*Híbridos.* Sistemas de identificação baseados em um único gene (nuclear ou mitocondrial) não garantem a correta identificação de híbridos (indivíduos cujo pai e mãe pertencem a espécies diferentes). Tais espécimens podem ser incorretamente identificados por morfologia também.

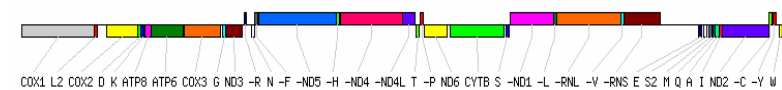
*Pseudogenes nucleares.* Pseudogenes (cópias inativas de genes que geralmente possuem várias mutações e/ou deleções) podem complicar a identificação baseada em genes mitocondriais e nucleares. Nos grupos já estudados, os pseudogenes não limitariam o uso do código de barras mitocondrial.

#### 5. Por que selecionar uma seqüência de código de barras dentro de um gene?

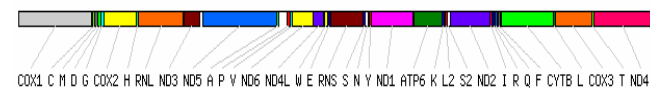
Com algumas exceções, as mitocôndrias animais contêm o mesmo conjunto de genes: 13 que codificam proteínas, dois RNA ribossomais e 22 RNA transportadores. Apesar do arranjo e da polaridade (fita leve ou pesada) dos genes diferirem bastante entre os filos, seqüências de diversos organismos podem ser facilmente comparadas desde que o código de barras esteja limitado a um gene. Limitar-se a um gene também facilita o desenvolvimento de técnicas abrangentes para a obtenção de seqüências de vários organismos.

A organização do genoma mitocondrial varia entre os animais. Por exemplo, os genomas da abelha e de um nematóide são apresentados. Seus arranjos gênicos diferem em 37 pontos de quebra. Assim, não é possível usar seqüências que ultrapassam os limites dos genes.

*Apis mellifera* Abelha



*Necator ameicanus* Nematóide

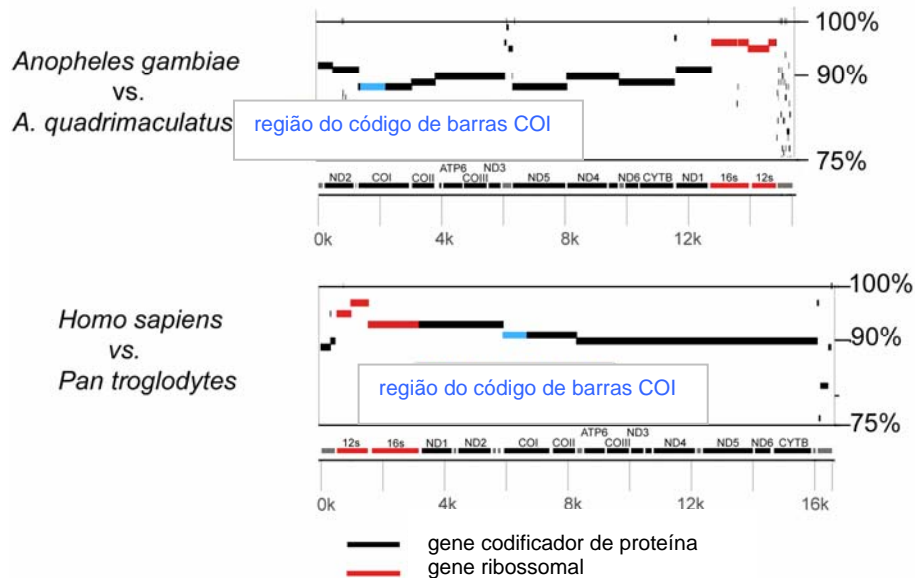


#### 6. Por que padronizar COI para animais?

Os genes codificadores de proteína mitocondriais contêm maior variabilidade do que os genes ribossomais e, portanto, seriam melhores para distinguir espécies próximas. Além disso, a comparação de seqüências de genes codificadores de proteínas é mais fácil, pois elas geralmente não apresentam inserções ou deleções, que são comuns em genes ribossomais.

Gráfico de porcentagem de identidade de genomas mitocondriais completos.

Os genes codificadores de proteínas geralmente apresentam mais diferenças entre espécies do que os genes ribossomais.



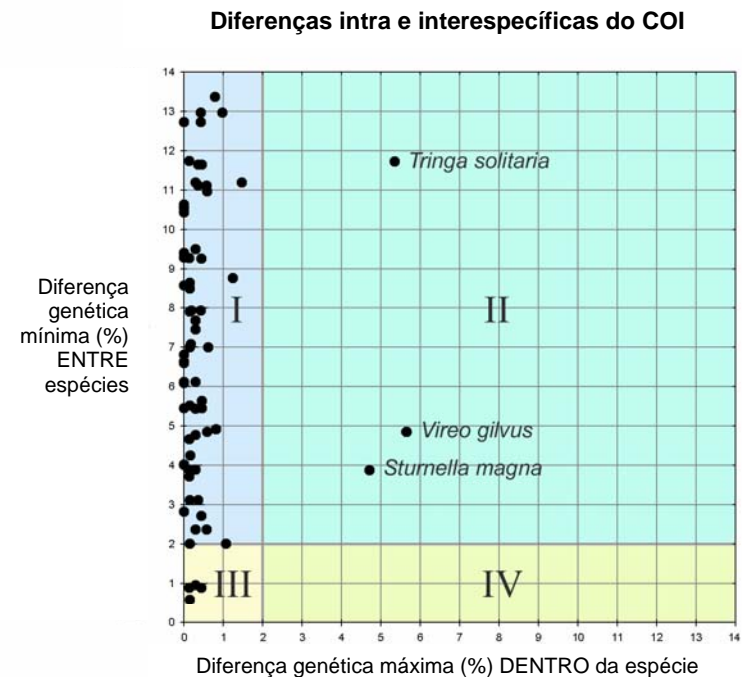
Entre as regiões candidatas de genes codificadores para proteína, a citocromo oxidase I (COI) possui variabilidade semelhante à encontrada em outros genes codificadores de proteína. O ganho em precisão ou custo no uso de outro segmento codificador de domínio protéico possivelmente seria pequeno, dada a alta similaridade geral entre essas regiões.

A região da COI que está rapidamente sendo aceita é a primeira metade do gene e possui 648 pares de bases, que é um tamanho facilmente processado com a tecnologia disponível e, portanto, barato. Os resultados indicam que esse código de barras COI é:

- 1) fácil de ser obtida para vários táxons usando um número pequeno de *primers*
- 2) é facilmente alinhada para comparação de seqüências
- 3) é informativo para distinguir espécies próximas de animais (vários invertebrados e vertebrados)

## 7. O que as diferenças do código de barras entre e dentro de espécies de animais estão apontando?

As seqüências de código de barras COI diferem muito mais entre do que dentro de espécies. Por exemplo, 260 espécies de aves norte-americanas apresentam, em média, diferenças entre espécies próximas 18 vezes maior do que as diferenças dentro das espécies. Ou seja, o código de barras COI consegue identificar a maioria dessas espécies. Há exceções, como espécies que se divergiram recentemente ou que hibridam frequentemente. Alternativamente, poucas diferenças no código de barras entre espécimens atribuídos a diferentes espécies podem indicar sinonímia, ou seja, uma única espécie incorretamente identificada como duas. Por outro lado, uma grande diferença no código de barras entre espécimens da mesma espécie indicaria a presença de mais de uma espécie no grupo descrito pela taxonomia corrente.



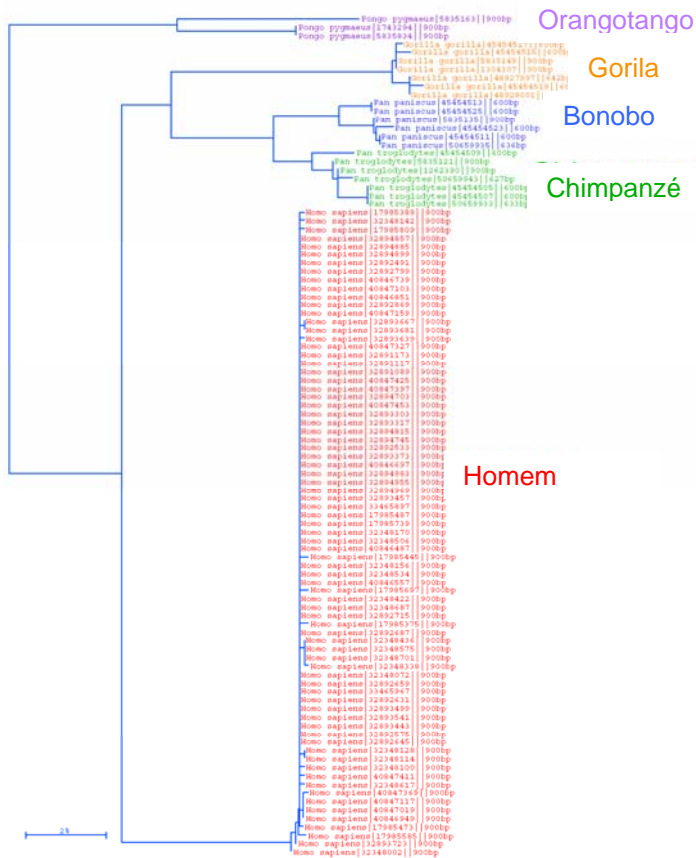
Resultados entre 73 espécies de aves da América do Norte. Os quadrantes representam diferentes categorias de espécies:

- I. congruente com a taxonomia vigente
- II. possíveis espécies crípticas (candidatas à separação taxonômica)
- III. divergência recente, hibridação, ou possível sinonímia
- IV. possível erro de identificação

## 8. E no ser humano?

O código de barras evidencia a unidade da espécie *Homo sapiens*. A comparação de seqüências de COI mostra que as pessoas são diferentes em apenas um ou dois pares de bases de um total de 648, enquanto diferimos dos chimpanzés em cerca de 60 sítios e dos gorilas em 70 sítios. Diferenças intraespecíficas grandes podem indicar a presença de espécies crípticas, como aconteceu com duas espécies de orangotango.

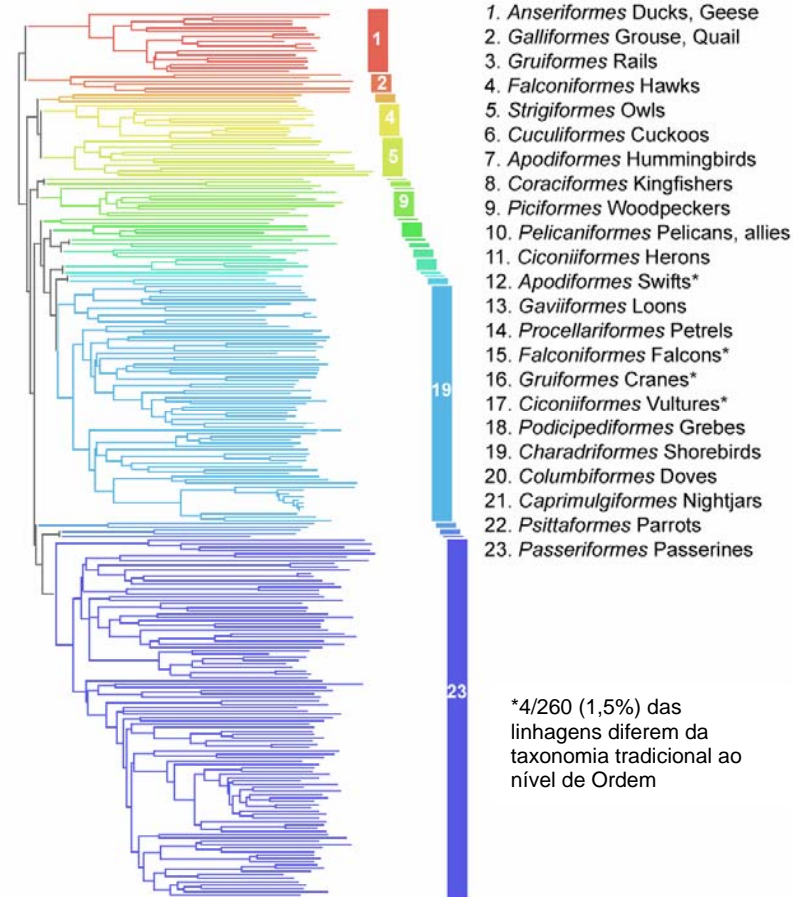
Árvore de *Neighbor-joining* baseada em distâncias genéticas do COI entre e dentro de 100 Hominidae



## 9. O código de barras pode auxiliar na compreensão da história de espécies de animais e de plantas?

Apesar de o objetivo do código de barras ser a identificação de espécimens ao nível de espécie, alguns pesquisadores apresentam os resultados das análises de código de barras em árvores, indicando distâncias evolutivas e relações entre os grupos. Há séculos os biólogos têm buscado obter uma árvore da vida ou filogenia que mostra a história das espécies. Esses estudos se beneficiam da análise de vários caracteres, especialmente ao longo de várias eras e grupos. Nos poucos casos analisados até o momento, as distâncias genéticas entre os códigos de barras são congruentes com o conhecimento gerado pela taxonomia tradicional, indicando que a biblioteca de códigos de barras pode auxiliar em estudos evolutivos.

Árvore de *Neighbor-joining* de código de barras de COI de 260 espécies de aves norte-americanas



## 10. Quem é responsável pelo avanço do código de barras da vida?

O Consórcio do Código de Barras da Vida (abreviação no inglês CBOL) é uma colaboração internacional entre museus de história natural, herbários, coleções biológicas e instituições que realizam inventários biológicos, bem como de especialistas em genômica, taxonomia, eletrônica e ciência da computação da Academia e de empresas comerciais. O objetivo do CBOL é acelerar a compilação de códigos de barras de DNA de espécies bem conhecidas ou recentemente descritas de animais e de plantas, estabelecer um banco público de seqüências correlacionadas a espécimens e promover o desenvolvimento de um equipamento portátil para identificação por meio do código de barras da vida. Mais informações são encontradas em:

<http://barcoding.si.edu>      <http://www.barcodinglife.org>  
<http://phe.rockefeller.edu/BarcodeConference/index.html>

### FONTES

#### *Espécies conhecidas de não descritas*

Blaxter M. 2003. Counting angels with DNA. *Nature* 421: 122-124.  
Tudge C. 2000. *The Variety of Life*. 684pp. Oxford University Press.

#### *Por que padronizar?*

Hebert PDN, Cywinska A, Ball SL, deWaard, JR. 2003. Biological identifications through DNA barcodes. *Proc Royal Soc Lond B* 270: 313-322.  
Janzen DH. 2004. Now is the time. *Phil Trans Royal Soc Lond B* 359:731-732.

#### *DNA mitocondrial na identificação de espécies*

Avice JC, Walker D. 1999. Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. *Proc Natl Acad Sci* 96:992-995.  
Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M. 2003. Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus *Homo*. *Proc Natl Acad Sci USA* 100: 7181-7188.

#### *Composição do genoma mitocondrial e ordem dos genes*

Jameson D, Gibson AP, Hudelot C, Higgs PG. 2003. ORGe: a relational database for comparative analysis of mitochondrial genomes. *Nucl Acids Res* 31: 202-206.

#### *Por que usar o COI em espécies animais?*

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3: 294-299.  
Hebert PDN, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: cytochrome *c* oxidase subunit I divergences among closely related species. *Proc R Soc Lond B* 270: S596-S599.  
Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W. 2000. PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res* 10: 577-586.

#### *Diferenças entre e intra espécies*

Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM. 2004. Identification of birds through DNA barcodes. *PLoS Biol* 2:1657-1663.  
Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrapes fulgerator*. *Proc Natl Acad Sci USA* 101:14812-14817.  
Non-human primate COI barcode sequences provided by Coriell Institute for Medical Research (<http://www.IPBIR.org>)

## Uma publicação ilustrada prévia disponível DEZ RAZÕES PARA TER O CÓDIGO DE BARRAS DA VIDA:

- 1. Usa fragmentos.** O código de barras pode identificar uma espécie a partir de pequenos fragmentos, incluindo material contaminante de animal ou de planta em alimentos processados e produtos morfológicamente não identificáveis derivados de espécies protegidas ou regulamentadas.
- 2. Não depende do estágio de vida.** O código de barras permite identificar a espécie em suas várias formas, de ovos a sementes, de larvas a plântulas, de adultos a flores.
- 3. Revela semelhantes.** O código de barras pode distinguir espécies semelhantes, revelando organismos perigosos que se passam por inofensivos e permitindo uma visão mais precisa da biodiversidade.
- 4. Reduz ambiguidades.** O código de barras permite realizar a identificação inequívoca de espécies, complementando as gradações de palavras, formas e cores.
- 5. Alavanca o conhecimento.** Cientistas podem usar o código de barras para acelerar a identificação de espécies descritas e reconhecer espécies novas.
- 6. Acesso democrático.** Uma biblioteca de código de barras permitirá que mais pessoas possam identificar espécies.
- 7. Abre a perspectiva de um guia de campo eletrônico.** O código de barras conecta a identificação biológica às fronteiras do sequenciamento do DNA, da eletrônica e da ciência da informação, formando a base para a obtenção de um aparelho portátil para a identificação de espécies.
- 8. Novas folhas brotando da árvore da vida.** Semelhanças e diferenças detectadas pelo código de barras entre as 10 milhões de espécies estimadas de animais e plantas auxiliarão a mostrar onde as folhas se encontram na árvore da vida.
- 9. Mostra o valor das coleções.** A compilação da biblioteca do código de barras se inicia com vários milhões de espécimens em museus, herbários, zoológicos, jardins e outras coleções biológicas, destacando seus esforços de preservação e de estudo da biodiversidade da Terra.
- 10. Acelera a obtenção da enciclopédia da vida.** A biblioteca do código de barras ligado a espécimens identificados aumentará o acesso público ao conhecimento biológico, ajudando a criar uma enciclopédia da vida na Terra *on line*.

<http://phe.rockefeller.edu/barcode>