



Capítulo 7

Diversidade Biológica

Free Delta: um sistema de software livre para o processamento de descrições taxonômicas

Mauro José CAVALCANTI

Ecoinformatics Studio. Caixa Postal 46521 CEP 20551-970 – Rio de Janeiro – RJ – Brasil.
E-mail: maurobio@gmail.com

Edinaldo Nelson dos SANTOS-SILVA

Coordenação de Pesquisas em Biologia Aquática/INPA. Av. André Araújo 2936, Aleixo
CEP 69060-001 – Manaus – AM – Brasil. E-mail: nelson@inpa.gov.br

RESUMO - O sistema DELTA (*Description Language for Taxonomy*) consiste de um formato flexível para a codificação de descrições taxonômicas e um conjunto de programas para o manejo e organização da informação taxonômica, como construção de chaves dicotômicas, produção de descrições em linguagem natural e identificação interativa. Todavia, o sistema DELTA original foi baseado em um modelo de software proprietário, que coloca limites à utilização e expansão do sistema. O Projeto Free DELTA é uma alternativa de software livre e código aberto para um sistema genérico de processamento de descrições taxonômicas, oferecendo uma ferramenta de suporte acessível para inventários sistemáticos de biodiversidade.

PALAVRAS-CHAVE: identificação biológica, chaves taxonômicas, bioinformática, inventários de biodiversidade, DELTA.

INTRODUÇÃO

A identificação de organismos representa um dos passos mais importantes em qualquer estudo no âmbito das Ciências Biológicas. Áreas de grande interesse prático, como o combate às pragas agrícolas, dependem fortemente da identificação correta dos organismos envolvidos (insetos, fungos, etc.), a fim de que medidas eficazes de controle possam ser adotadas. A identificação precisa de organismos também é necessária para que seja possível utilizar-se a classificação taxonômica como um sistema eficiente de armazenamento e recuperação de informações (Pankhurst 1991; Edwards & Morse 1995).

Via de regra, a identificação de espécimes biológicos é feita por profissionais sediados em museus, herbários e universidades, os quais lançam mão de conhecimento sistemático para obter uma identificação correta, a partir do exame detalhado dos organismos e comparações destes com exemplares de coleção, ilustrações e descrições. Tal processo baseia-se muitas vezes na heurística do pesquisador, podendo um espécime ser reconhecido de imediato pelo especialista, sem nenhum procedimento mental explícito – o que requer, no entanto, uma quantidade razoável de experiência acumulada e perícia (Waldrop 1984).

Especialistas com tal nível de capacitação, usualmente existem em pequeno número (Gaston & May 1992) e, portanto, são frequentemente muito demandados. Em problemas práticos, como a identificação de uma praga agrícola, onde a precisão e a rapidez são vitais, a necessidade do envio de espécimes ao pesquisador em um grande centro pode acarretar sérios prejuízos à lavoura, devido ao tempo requerido para o envio e devolução do material identificado.

A identificação de organismos tornou-se mais fácil, rápida e acessível ao não especialista a partir da introdução das *chaves taxonômicas*, certamente o método de identificação biológica mais conhecido e utilizado desde o Século XVIII (Metcalf 1954; Papavero & Martins 1983). Uma chave taxonômica é um artifício usado para selecionar caracteres de

um determinado grupo de organismos de modo que, por uma série de escolhas alternativas, seja possível chegar à identificação de um espécime desconhecido. Chaves taxonômicas são frequentemente dicotômicas: para identificar um espécime, o usuário da chave escolhe, do primeiro par de opções contrapostas, aquela que for verdadeira para o organismo em questão. A opção escolhida pode ser seguida por um nome que identifica o organismo, ou por um número, que dirige o usuário a outro par de opções, continuando o processo até que uma identificação seja obtida.

Todavia, chaves taxonômicas são tradicionalmente elaboradas por taxonomistas humanos, que empregam procedimentos subjetivos na construção de uma chave para um determinado grupo de organismos e, com frequência, muito diferentes de um especialista para outro. Assim, ainda que as chaves taxonômicas tenham tornado possível ao não-especialista identificar organismos com relativa facilidade e rapidez, permaneceram incapazes de garantir flexibilidade e precisão nas identificações obtidas, sobretudo quando os caracteres necessários para uma identificação correta são difíceis de ver ou de reconhecer pelo usuário inexperiente.

A subjetividade envolvida no processo de construção de uma chave conduziu às primeiras tentativas de desenvolvimento dos sistemas de identificação por computador, capazes de construir chaves (ou identificar organismos interativamente) segundo procedimentos bem definidos, a partir dos dados fornecidos pelo taxonomista. Além de possibilitar, via de regra, uma identificação mais rápida pelo fato de serem otimizadas, as chaves produzidas por computador mostram-se muito mais flexíveis, permitindo a inclusão de novos caracteres ou táxons e revisões do grupo; ademais, tornam possível a construção, a partir da mesma base de dados, de diferentes tipos de chaves, adequadas ao uso em laboratório ou em campo, específicas para a fauna ou a flora de certas localidades ou determinados estágios do ciclo de vida dos organismos (Pankhurst 1991; Edwards & Morse 1995).



O SISTEMA DELTA

Uma das maiores contribuições no sentido de padronizar o processo de aquisição e processamento da informação taxonômica por computador foi a introdução do sistema DELTA (*DEscription Language for TAXonomy*) por Dallwitz (1980). Este sistema consiste de um formato flexível para a codificação da informação taxonômica descritiva e um conjunto de programas para o processamento de dados taxonômicos (Dallwitz 1974, 1980). O desenvolvimento do sistema DELTA começou em meados dos anos 70 na CSIRO Division of Entomology, Austrália, inicialmente para computadores de grande porte e depois adaptado para microcomputadores da linha IBM-PC. Este sistema oferece uma série de facilidades para o manejo e a organização da informação taxonômica. Especialmente úteis são a possibilidade de construção de chaves de identificação de diversos tipos, a produção de descrições em linguagem natural e a preparação de matrizes de dados para análises fenéticas e cladísticas, além da identificação interativa e recuperação de informações. O sistema DELTA representou o auge do desenvolvimento dos sistemas de identificação por computador, oferecendo não apenas algoritmos de identificação mais eficientes, mas também um formato prático e padronizado para a codificação de descrições taxonômicas (Askevold & O'Brien 1994).

O formato DELTA aceita todos os tipos de caracteres, tanto qualitativos (duplo-estado e multiestado, ordenados ou não-ordenados) quanto quantitativos (contínuos e descontínuos) e inclui instruções (diretivas) para controlar o processamento dos dados pelos diversos programas componentes do sistema. Uma vez codificados no formato DELTA, os mesmos dados podem ser utilizados para produzir descrições e chaves de identificação com propósitos gerais ou específicos, omitindo parte dos caracteres, ou dando ênfase a alguns deles, o que permite a construção, a partir da mesma base de dados, de diferentes tipos de chaves, adequadas ao uso em laboratório ou em campo, específicas para a fauna ou a flora de certas localidades ou determi-

nados estágios do ciclo de vida dos organismos. Isto faz do sistema DELTA uma valiosa ferramenta para inventários sistemáticos de biodiversidade (Allkin *et al.* 1992; Sharkey 2001) e levantamentos ecológicos de campo para estudos de impacto ambiental (Ellis 1988).

Allkin *et al.* (1992) demonstraram a utilidade do sistema DELTA para o manejo de dados e a produção de chaves e descrições taxonômicas no decorrer do inventário florístico de Veracruz, México. Sharkey (2001) aplicou o sistema DELTA no inventário exaustivo da biodiversidade do Parque Nacional das Montanhas Great Smoky, Estados Unidos. Ellis (1988) considerou o sistema DELTA essencial para implementar o “controle de qualidade” das identificações taxonômicas efetuadas no âmbito de levantamentos ecológicos para avaliações de impacto ambiental.

O programa INTKEY (Dallwitz 1993; Dallwitz *et al.* 1998), para identificação interativa e recuperação de informações, que faz parte do sistema DELTA, permite ao usuário acessar o conteúdo de um banco de dados taxonômico a partir de qualquer caráter e na ordem ou combinação em que desejar, iniciando tantas vezes quantas necessárias até a determinação de um táxon ser confirmada. O INTKEY permite também o armazenamento de ilustrações digitalizadas associadas às descrições dos caracteres para guiar o usuário ao longo da chave. Várias bases de dados taxonômicos têm sido criadas com o sistema DELTA, compreendendo desde gêneros de gramíneas e leguminosas até espécies de corais, formigas e besouros. Muitas destas bases de dados incluem imagens digitalizadas dos organismos e caracteres e estão disponíveis na Internet (<http://www.delta-intkey.com>).

Outros programas para o processamento de dados taxonômicos codificados em formato DELTA foram desenvolvidos na Inglaterra por Richard Pankhurst (PANKEY: Pankhurst 1986), na Holanda por Eric Gouda (TAXASOFT: Gouda 1996), na Alemanha por Gregor Hagedorn (DELTA Access: Hagedorn & Rambold 2000), na Austrália por Jason Nunn (MONTANUS) e Mike Choo (DELIA: Chapman & Choo 1996) e na Espanha por Antonio Garcia-

Valdecasas (EDEL: Bello *et al.* 1996). Ambientes interativos para a utilização integrada dos vários programas que compõem o sistema DELTA original foram desenvolvidos na Austrália por Nicholas Lander (DELTA Menu System: Lander 1993) e, no Brasil, pelo autor senior (DIANA: Cavalcanti 1996).

Todavia, a despeito de sua flexibilidade e eficiência na codificação de descrições taxonômicas, a maior parte dos programas desenvolvidos para o processamento de dados no formato DELTA apresentam algumas desvantagens, a saber: (1) não disponibilizam o código-fonte para inspeção e modificação, de acordo com as necessidades de cada usuário ou comunidade de usuários; (2) são limitados a um sistema operacional proprietário (Microsoft Windows).

O PROJETO FREE DELTA

O projeto Free DELTA (<http://freedelta.sourceforge.net>) foi iniciado em abril de 2000, com o objetivo de criar um sistema completo, de código livre e aberto, compatível com o sistema operacional Linux, para o processamento de descrições taxonômicas codificadas no formato DELTA, em seguida à decisão da CSIRO Division of Entomology de interromper o suporte ao desenvolvimento dos programas do sistema DELTA original.

Em maio de 2005, o projeto Free DELTA foi registrado no SourceForge, o maior repositório mundial de desenvolvimento de software de código aberto disponível na Internet, onde o site do Free DELTA está agora hospedado em bases permanentes.

Em julho de 2008, a iniciativa Free DELTA NG (“New Generation”) foi lançada, para promover o desenvolvimento e intercâmbio de ferramentas de software livre e de código aberto para DELTA, inteiramente baseadas em tecnologias da Internet, portanto acessíveis de qualquer lugar, de acordo com o princípio da “computação em nuvem”. A iniciativa Free DELTA NG também define as condições que uma dada ferramenta de software deve atender a fim de ser considerada como um programa Free DELTA program (ver abaixo).

Em sua versão inicial, o Free DELTA consistia apenas de uma biblioteca de classes em linguagem

C++ para a leitura de arquivos no formato DELTA, escrita por Denis Ziegler, com contribuições adicionais de Guillaume Rousse e Bastiaan Wakkie.

A partir da versão “NG”, o sistema Free DELTA passou a ser composto por outros módulos e correspondentes funções:

PyDELTA (<http://freedelta.sourceforge.net/pydelta>): biblioteca de classes em linguagem Python para a manipulação de arquivos de dados no formato DELTA, desenvolvida por Mauro J. Cavalcanti & Thomas Kluyver.

WebDelta (<http://freedelta.sourceforge.net/webdelta>): ferramenta em linguagem Perl para edição de dados em formato DELTA através de um navegador Web, escrito por Claudio Rivetti e Riccardo Percudani.

NaviKey (<http://www.navikey.net>): mini-aplicativo (*applet*) Java para identificação interativa a partir de dados taxonômicos codificados no formato DELTA, desenvolvido por Michael Bartley, Noel Cross, Dieter Neubacher e Gerhard Rambold.

A biblioteca de programação do Free DELTA fornece um conjunto padronizado, versátil e extensível de rotinas para a manipulação de dados no formato DELTA. A partir da versão “NG” do Free DELTA, esta biblioteca passou a ser desenvolvida em Python (<http://www.python.org>), uma linguagem de programação interpretada, orientada a objetos, com uma sintaxe clara e concisa, adequada ao desenvolvimento de aplicações para acesso a bancos de dados, Internet e computação científica (Beazley 2000; Bassi 2007; Oliphant 2007). Este conjunto de rotinas foi projetado para fornecer o suporte básico para todos os programas do Free DELTA; além disso, as rotinas podem ser prontamente incorporadas em programas aplicativos individuais ou sistemas de bancos de dados que sejam criados por diferentes desenvolvedores. Estas rotinas são distribuídas como pacotes de código-fonte livre e aberto. Assim, um usuário que necessite de um programa para manejo de dados taxonômicos diferente dos utilitários do Free DELTA, está apto a desenvolvê-lo por si mesmo, ou contratar um desenvolvedor individual ou empresa de desenvolvimento para desenvolvê-lo, ao mesmo tempo man-



tendo os benefícios do formato padronizado para a codificação de dados taxonômicos oferecidos pelo DELTA. A disponibilidade do código-fonte completo também permite ao usuários do Free DELTA adaptar estas rotinas de acordo com suas necessidades específicas. Espera-se que isto contribua para expandir a comunidade de usuários do DELTA, bem como fortalecer o papel do DELTA como um sistema padronizado de codificação e intercâmbio de dados taxonômicos.

O Free DELTA é capaz de processar dados codificados em DELTA, mas não é idêntico ao sistema DELTA original. Estão sendo feitos vários aperfeiçoamentos, com base na experiência com outros sistemas de computação taxonômica, em particular um melhor tratamento de dados quantitativos, mais procedimentos analíticos e de conversão de formatos e uma interface gráfica de usuário mais simples e flexível.

O Free DELTA tem como características básicas:

- 1) a disponibilidade do código-fonte completo, assegurando a cada usuário ou comunidade de usuários a liberdade para modificar os programas e adaptá-los às suas necessidades específicas;
- 2) a possibilidade de produzir análises quantitativas de dados descritivos e apresentá-las sob a forma de tabelas e gráficos;
- 3) a capacidade de exportação de quaisquer subconjuntos de dados em vários formatos padronizados de intercâmbio de dados biológicos descritivos, para utilização em outros programas.

Desenvolvedores de software e biólogos taxonomistas são sempre bem-vindos ao projeto Free DELTA. Programadores individuais podem contribuir escrevendo um substituto compatível para qualquer programa do sistema DELTA original, doando-o ao projeto Free DELTA com o código-fonte completo e documentação, nos termos da mesma licença de uso e distribuição adotada pelo projeto Free DELTA.

Para ser incluída no Free DELTA, espera-se que uma ferramenta de software atenda às seguintes condições:

Deve ser livre, em ambos os sentidos em que esta palavra é usada no mundo da computação (isto é, “de código aberto” e “gratuita”).

Deve ser multi-plataforma (isto é, não restrita a sistemas operacionais específicos, especialmente sistemas proprietários)

Deve ser totalmente compatível com o formato DELTA original (isto é, não pretendendo substituir ou fornecer extensões *ad hoc* para este formato).

Os programas do Free DELTA devem rodar com pouca (ou nenhuma) modificação em vários sistemas operacionais. O código-fonte completo é distribuído nos termos da Licença Pública GNU (<http://www.gnu.org>). Módulos executáveis pré-compilados dos programas do Free DELTA (conversão de formatos, construção de chaves de identificação, identificação interativa e edição de dados taxonômicos descritivos) também são disponibilizados, para computadores pessoais com GNU/Linux, MS-Windows e Mac OS X. O sistema também inclui documentação técnica e de usuário (na forma de manuais e tutoriais), que fornecem toda a informação necessária para usar e modificar os programas.

No âmbito do Projeto Biotupé, o sistema Free DELTA será usado para produzir guias de identificação (tanto em formato eletrônico quanto na forma de manuais impressos) a partir de descrições das espécies coletadas pelos diversos projetos envolvidos no inventário da biodiversidade da RDS do Tupé. Assim, o Projeto Biotupé espera contribuir para a implementação de uma metodologia eficiente e acessível para a elaboração automatizada de guias de biodiversidade, que permitam a usuários tecnicamente habilitados, mas não necessariamente especializados em botânica ou zoologia, identificarem as espécies de plantas e animais da área.

Como se trata de produto inêdito em sua atualidade e funcionalidade, o *Free DELTA* ampliará a autonomia tecnológica e científica nacional na área de bioinformática e permitirá que grupos de pesquisa ou pesquisadores individuais envolvidos em estudos de biodiversidade tenham pleno o acesso às modernas tecnologias de manejo de dados taxonômicos, contribuindo para uma aplicação mais eficiente e efetiva do conhecimento biológico em

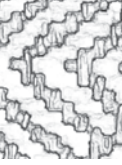
esforços de conservação e utilização sustentável das espécies vivas, especialmente no contexto do sistema megadiverso da Amazônia, com benefícios para toda a sociedade brasileira.

AGRADECIMENTOS

Aos Drs. Mike Dallwitz (CSIRO Division of Entomology, Canberra) e Richard Pankhurst (Royal Botanic Gardens, Edinburgh) pelo envio de publicações; aos Drs. Tatiana Paleo Konno (Universidade Federal do Rio de Janeiro) e Jeffrey Jon Shaw (ICB, Universidade de São Paulo) pela leitura crítica do manuscrito e valiosas sugestões.

BIBLIOGRAFIA CITADA

- Allkin, R., Moreno, N. P., Gama Campillo, L. & Mejia, T. 1992. Multiple uses for computer-stored taxonomic descriptions: keys for Veracruz. *Taxon*, 41: 413-435.
- Askevold, I. S. & O'Brien, C. W. 1994. DELTA, an invaluable computer program for generation of taxonomic monographs. *Annals of the Entomological of America*, 87: 1-16.
- Bassi, S. 2007. A primer on Python for life science researchers. *PLoS Computational Biology*, 3: 2052-2057.
- Beazley, D. M. 2000. Scientific computing with Python. In: Manset, N., Veillet, C. & Crabtree, D. (eds.) *Astronomical Data Analysis Software and Systems IX*. ASP Conference Series, Vol. 216, p. 49-58.
- Bello, E., Valdecasas, A. G.- & Elvira, J. R. 1996. A Spanish and English DELTA Editor: EDEL 95. *DELTA Newsletter*, 12: 24-26.
- Cavalcanti, M. J. 1996. Designing a graphical user interface for DELTA: some considerations. *DELTA Newsletter*, 12: 2-4.
- Cavalcanti, M. J. 1998. Aplicações de Sistemas Gerenciadores de Bancos de Dados em Ecologia e Sistemática. *Acta Biologica Leopoldensia*, 20: 5-20.
- Chapman, A. R. & Choo, M. 1996. Institutional DELTA databases: a case study. *DELTA Newsletter*, 12: 14-15.
- Dallwitz, M. J. 1974. A flexible computer program for generating identification keys. *Systematic Zoology*, 23: 50-57.
- Dallwitz, M. J. 1980. A general system for coding taxonomic descriptions. *Taxon*, 29: 41-46.
- Dallwitz, M. J. 1993. DELTA and INTKEY. In: Fortuner, R. (Ed.). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. The Johns Hopkins University Press, Baltimore, Maryland, p. 287-296.
- Dallwitz, M. J., Paine, T. A. & Zurcher, E. J. 1998. Interactive keys. In: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (Eds.). *Information Technology, Plant Pathology and Biodiversity*. CAB International, Wallingford, p. 201-212.
- Edwards, M. & Morse, D. R. 1995. The potential for computer-aided identification in biodiversity research. *Trends in Ecology and Evolution*, 10: 153-158.
- Ellis, D. V. 1988. Quality control of biological surveys. *Marine Pollution Bulletin*, 19: 506-512.
- Gaston, K. J. & May, R. M. 1992. Taxonomy of taxonomists. *Nature*, 356: 281-282.
- Gouda, E. J. 1996. TAXASOFT DELTA Editor. *DELTA Newsletter*, 12: 12-14
- Hagedorn, G. & Rambold, G. 2000. A method to establish and revise descriptive data sets over the Internet. *Taxon*, 49: 517-528.
- Lander, N. S. 1993. DELTA Menu System for Windows. *DELTA Newsletter*, 8: 10-11.
- Metcalf, Z. P. 1954. The construction of keys. *Systematic Zoology*, 3: 38-45.
- Oliphant, T. E. 2007. Python in scientific computing. *Computing in Science & Engineering* 9: 10-20.
- Pankhurst, R. J. 1986. A package of computer programs for handling taxonomic databases. *Computer Applications in the Biosciences*, 2: 33-39.
- Pankhurst, R. J. 1991. *Practical Taxonomic Computing*. Cambridge University Press, Cambridge. 202p.
- Papavero, N. & Martins, U. R. 1983. Itens da publicação taxonômica. In: Papavero, N. (Ed.). *Fundamentos Práticos de Taxonomia Zoológica*:



- Coleções, Bibliografia, Nomenclatura. Museu Paraense Emílio Goeldi/Sociedade Brasileira de Zoologia, Belém, p. 73-112.
- Sharkey, M. J. 2001. The All Taxa Biological Inventory of the Great Smoky Mountains National Park. Florida Entomologist, 84: 556-564.
- Waldrop, M. M. 1984. The necessity of knowledge. Science, 223: 1279-1282.

