



Review

DNA barcode information for the sugar cane moth borer *Diatraea saccharalis*

J.P. Bravo, J.L.C. Silva, R.E.F. Munhoz and M.A. Fernandez

Departamento de Biologia Celular e Genética,
Universidade Estadual de Maringá, Maringá, PR, Brasil

Corresponding author: M.A. Fernandez
E-mail: aparecidafernandez@gmail.com

Genet. Mol. Res. 7 (3): 741-748 (2008)
Received May 8, 2008
Accepted July 4, 2008
Published August 14, 2008

ABSTRACT. We reviewed the use and relevance of barcodes for insect studies and investigated the barcode sequence of *Diatraea saccharalis*. This sequence has a high level of homology (99%) with the barcode sequence of the Crambidae (Lepidoptera). The sequence data can be used to construct relationships between species, allowing a multidisciplinary approach for taxonomy, which includes morphological, molecular and distribution data, all of which are essential for the understanding of biodiversity. The *D. saccharalis* barcode is a previously undescribed sequence that could be used to analyze Lepidoptera biology.

Key words: Cytochrome C Oxidase Subunit I; Mitochondrial sequence; Barcodes; Lepidoptera; *Diatraea saccharalis*

ANALYSIS OF DNA BARCODES

A typical metazoan mitochondrial DNA (mtDNA) genome is composed of a 14-39-kb double-stranded circular molecule. It encodes 13 protein-coding genes, two genes that encode ribosomal RNAs, and 22 genes that encode transfer RNAs (Wolstenholme, 1992). It also contains non-coding DNA rich in A-T sequences necessary for the initiation and regulation of transcription and replication (Boore and Brow, 1998).

Variations in the mtDNA sequence are among the most widely used genetic markers in animals, because it is a haploid genome, it is easily amplified from a variety of taxa, and sequencing can easily be performed without cloning. The high rate of evolution that occurs within the mitochondrial genome allows the pattern and timing of recent historical events to be deduced without extensive sequencing efforts (Hurst and Jiggins, 2005). mtDNA has been extensively used in studies of phylogenetics, phylogeography, the dynamics and structure of populations, and molecular evolution (Zhang and Hewitt, 1997).

Hebert et al. (2003) proposed that a universally accessible database of COI barcodes should be constructed. This approach utilizes a fragment of approximately 658 bp of the first half of the mitochondrial Cytochrome C Oxidase Subunit I gene, named COXI or COI. The use of a common DNA sequence, or a set of DNA sequences across a wide range of taxa with a uniform format for the submission, accession, and storage of tissues and information would greatly benefit our understanding of biodiversity (DeSalle, 2006).

The Consortium for the Barcodes of Life (CBOL) is formed by major natural history museums, universities, herbaria, and other organizations. The aim of this consortium is to establish the ambitious “The Barcodes of Life Initiative”, with the intent of using DNA barcodes to identify each of the estimated 10 million species on earth (Savolainen et al., 2005). The DNA segment used for the barcode is approximately 658 bp of the mitochondrial gene Cytochrome C Oxidase I, COI. This sequence is suitable as a central part of a global identification system because it can easily be amplified from variety of taxa, it is a haploid genome, it displays a maternal pattern of inheritance, and it has a high rate of evolution. The CBOL was launched in May 2004 and includes more than 150 organizations from 45 nations, including universities, departments of biology and molecular biology, natural history museums, and herbaria.

Currently, the efficacy of DNA barcoding is being assessed using tools established by the CBOL, as outlined in the Barcode of Life Data Systems (BOLD, 2008), website www.barcodinglife.org. It provides an integrated bioinformatics platform that supports all phases of the analytical pathway, from specimen collection to a highly validated barcode library (Figure 1; Ratnasingham and Hebert, 2007). BOLD was initially developed as an informatics workbench for a single, high volume DNA barcodes facility, and was used for the first major barcode project, which included birds, fish and Lepidoptera. The CBOL (2008) formally described 35,289 species with barcodes; it contained 335,714 barcode records.

BOLD employs several tools to identify data anomalies or low-quality records. All submitted sequences are first translated into amino acids and are compared against a Hidden Markov Model of the COI protein in order to check whether the sequences are actually derived from COI. Sequences that pass this test are then examined for stop codons and are compared

against a small suite of possible contaminants. If any potential errors are detected, the submitter is informed and the sequence is flagged (Ratnasingham and Hebert, 2007).

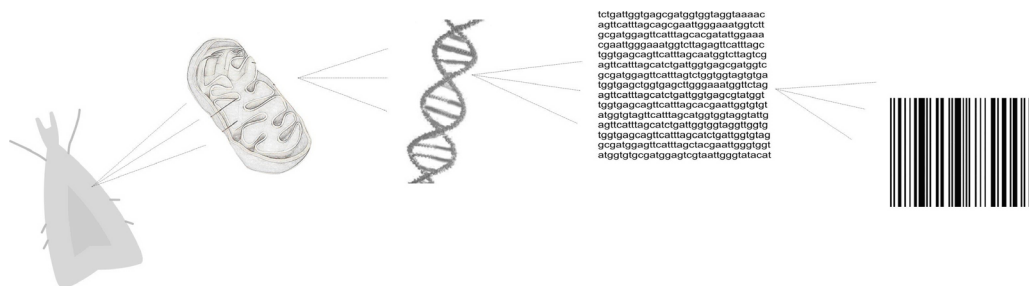


Figure 1. Barcoding invertebrates. Schematic representation of barcode sequence methodology.

Barcoding is emerging as a cost-effective standard for rapid species identification and has the potential to accelerate the discovery of new species and improve the quality of taxonomic information. It also makes this novel information readily available to non-taxonomists and research projects that are being conducted outside major collection centers (Miller, 2007).

The three main taxonomic applications that DNA barcoding has been previously used in are: 1) the identification of species previously defined by other criteria, including rapid identification, as well as linking specimens to established species that are unidentifiable by other means; 2) the description of new species by interpreting DNA diversity as an indicator of species diversity; 3) the definition of operational units for ecological studies (Rubinoff et al., 2006).

The applicability of COI barcodes to identified species has been demonstrated for a wide variety of organisms, including gastropods (Remigio and Hebert, 2003), tropical Lepidoptera (Hajibabaei et al., 2006a), blowflies (Nelson et al., 2007), tropical parasitoid flies (Smith et al., 2007), birds (Hebert et al., 2004) and fish (Ward et al., 2005). Traditional morphology-based taxonomic procedures are time-consuming and may not always be sufficient for identification to the species level; therefore, a multidisciplinary approach to taxonomy that includes morphological, molecular and distribution data is essential (Krzyszowski and Besansky, 2003).

DNA barcodes have emerged during a critical period for taxonomy. Economic development and increased international commerce are leading to higher extinction rates and to the introduction of invasive and pest species (Miller, 2007). Long-term research strategies are also required to address the deficiencies in existing taxonomic keys to deal with morphologically indistinct immature life stages, cryptic species and damaged specimens. An approach utilizing DNA barcodes would provide a very realistic, practical and flexible framework for species identification in the context of biosecurity (Armstrong and Ball, 2005). In Japan, four exotic insect species have become established each year on average over the last 50 years; 74% became economic pests (Kiritani, 1998).

Smaller fragments (100 bp) of the standard COI barcodes, “mini-barcodes”, have been found to be effective for species identification in samples where the DNA is degraded or in other situations where it is not possible to obtain a full-length barcode. The mini-barcodes generally can provide measures of sequence variability and divergence at similar levels to full barcodes, at both the intra-specific and intra-generic levels (Hajibabaei et al., 2006b).

Min and Hickey (2007) suggested that important components of the whole mitochondrial genome can be predicted with a high degree of accuracy from the short barcode sequence alone. These components include average nucleotide composition, patterns of strand asymmetry and a high frequency of codons that encode hydrophobic amino acids.

However, there are some issues to consider when using barcodes. There have been technical issues arising from the finding of nuclear integrations of mtDNA. mtDNA integrated into the nuclear genome may still amplify with conserved primers targeted at mtDNA, complicating or confounding analysis (Bensasson et al., 2001). Many arthropods carry microorganisms inside their cells, and females may transmit these microorganisms to their progeny. Factors such as interspecific hybridization and infection by maternally transmitted endosymbionts, such as *Wolbachia*, are now known to cause mitochondrial gene flow between biological species (Hurst and Jiggins, 2005). The groups created using mtDNA can differ from the true species cluster and may also confound interpretation and attempts to reconstruct the phylogeography of a species (Hurst and Jiggins, 2005). Heteroplasmy is also a potential problem for mtDNA analysis. Heteroplasmy is the existence of different mitochondrial haplotypes within individuals; this mitochondrial variability includes both sequence variability and length heteroplasmies due to insertions or deletions. The phenomenon of indels has not previously been addressed by proponents of barcodes (Rubinoff et al., 2006).

Ideally, an appropriate marker for barcoding species should display a high level of interspecific variability (to allow discrimination between closely related species); it should also have low levels of intraspecific variability (to allow specimens to be accurately assigned to species) (Rach et al., 2008).

Cywinska et al. (2006) analyzed sequence variation in the barcode region of the COI gene in order to test its usefulness for the identification of 37 species of Canadian mosquitoes (Diptera, Culicidae). Specimens of a single species formed barcode clusters with tight cohesion that were usually clearly distinct from those of related species. Min and Hickey (2007) studied the application of barcodes for the classification and phylogenetic reconstruction of unknown fungal species. They used 31 fungal species; including 27 Ascomycota, three Basidiomycota and one Chytridiomycota (outgroup). They found that short DNA barcodes (600 bp) can separate all 31 fungal species; these results were confirmed further with a phylogenetic tree. The COI barcodes for 260 species of North American birds allowed the identification of four potentially novel species, suggesting that a global survey using this method may lead to the recognition of many additional bird species (Hebert et al., 2004).

ANALYSIS OF LEPIDOPTERA USING DNA BARCODES

The consortium established the “All-Leps Barcodes of Life” project because the Lepidoptera are the second most diverse order of insects. There are about 180,000 known species, and it is likely that there are another 300,000 species awaiting description. The initiative involves campaigns on three geographic scales: Global (Geometridae, Saturniidae and Spingidae), Continental (North America and Australia) and Regional [Great Smoky Mountains National Park (USA) and Area de Conservación Guanacaste]. All-Leps Barcodes of Life, 2008, displayed 9698 barcoded Lepidoptera species.

Developing a DNA barcode system for individual species requires adequate initial taxonomic identification and the ability to retain intact specimens for future morphological analysis.

Once a division has been identified, returning to such material may yield reliable characteristics that may have previously been regarded as morphological variation within a species.

Studies of community structure, food web dynamics, biodiversity, and biomonitoring also depend upon the accuracy of species discrimination and identification (Ball et al., 2005). These DNA sequence-based hypotheses are then open for testing and may provide the stimulus and starting point for further taxonomic revision of a particular group (DeSalle, 2006).

Published studies that provide the basis for the barcode system may be biased by exceptional situations. The COI barcodes distinguish more than 95% of species; however, some groups are in need of taxonomic revision, and further investigations on many vertebrate and invertebrate groups are required (Ward et al., 2005; Hajibabaei et al., 2006b).

Hajibabaei et al. (2006a) obtained COI sequences from 4260 adults of morphologically defined species of tropical Lepidoptera (hesperiids, sphingids and saturniids) from Area de Conservación Guanacaste in northwest Costa Rica. Most of the species exhibited low levels of COI sequence variation, whereas some presented sequence diversity that rivaled levels found between very similar species. Of the 521 species that they examined, 97.9% were unambiguously identified, suggesting that DNA barcoding can be an effective tool for species recognition in tropical regions.

DIATRAEA SACCHARALIS BARCODE

The moth borers are a diverse group of Lepidoptera, constituted primarily of noctuids and pyraloids; they are important since they are major pests of sugarcane throughout the world (Lange et al., 2004). Separation of the pyralids from the crambines is one of the more contentious issues in Lepidopteran phylogenetics. The more conservative view places all pyraloid subfamilies in one family, the Pyralidae (Fletcher and Nye, 1984; Shaffer et al., 1996; Holloway et al., 2001). In 1925, Börner first noted that there was a distinct division within this group; he split them into the Pyraliformes and Crambidiformes. Minet (1985) refined this concept, and placed the pyraloid subfamilies in either the Pyralidae or the Crambidae, depending on the presence or absence of a praecinctorium and whether the tympanal organs were medially approximated or well separated.

Lange et al. (2004) reported on the phylogeny of 26 species of sugarcane moth borers (Lepidoptera: Noctuidae and Pyraloidea) using mitochondrial partial gene sequences of COII and 16S. The genus *Diatraea* is monophyletic; but in their study, *Diatraea* resolved into two main groups, the first contains *centrella*, *crambidoides* and *grandiosella* and the other group includes *busckella*, *rosa* and *saccharalis*.

Barcodes may provide a useful tool to resolve this taxonomy problem. The family Crambidae, subfamily Crambinae, has 1416 species with barcodes. The genus *Diatraea* has 13 barcode sequences, two *Diatraea crambinoides* sequences, and 11 *Diatraea evanescens* sequences, but these sequences are not available in the public domain.

Our research group recently described the *Diatraea saccharalis* mitochondrial control region, CR (Bravo et al., 2008). Sequence analysis demonstrated that this region of the *D. saccharalis* mitochondrial genome has high similarity with that of the Lepidopteran *Cydia pomonella*; however, these results did not clarify the taxonomic problem posed by *D. saccharalis*.

The development of *D. saccharalis* mtDNA barcodes would be useful. The first problem that we encountered was difficulty in amplifying the COI sequence of this Lepidoptera.

The primers, which enable amplification of the *D. saccharalis* COI, were originally developed for use in the nematode *Toxocara canis* (Sato et al., 2005). The sequence of this amplified product of 424 nucleotides displayed a higher homology (99%) with the Crambidae. The ClustalW alignment with some of these sequences exhibited scores between 88 and 84% (Figure 2).

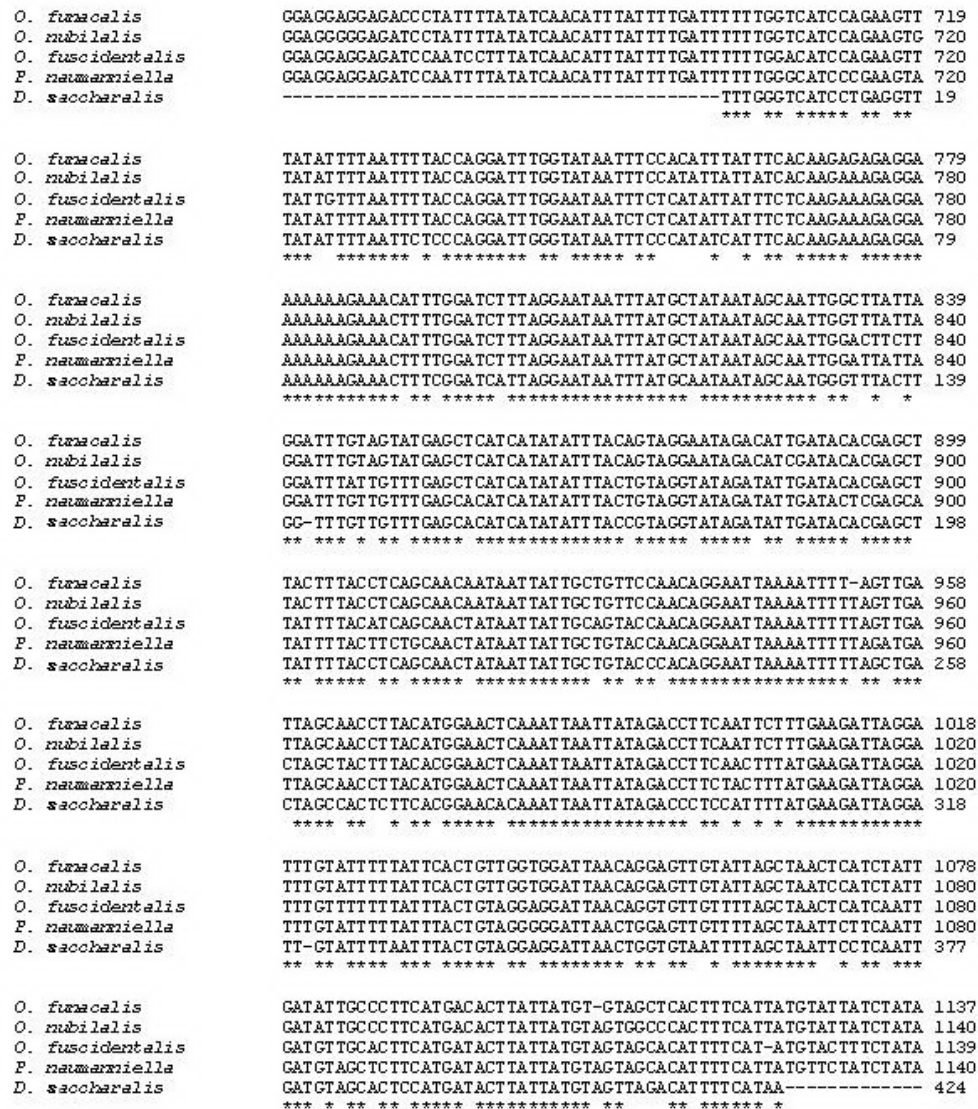


Figure 2. ClustalW analysis from mtDNA COI sequences of *Ostrinia funacalis* (NC003368), *Ostrinia nubilalis* (NC003367), *Omphisca fuscidentalis* (DQ523228), *Paracymoriza naumanniella* (AJ852523), and *Diatraea saccharalis*.

ACKNOWLEDGMENTS

We thank Valmir Peron and Marli Licero Schuete Silva for their dedicated technical assistance and the Universidade Estadual de Maringá (COMCAP laboratories) for facilities. Research supported by grants from FINEP/Fundação Araucária, Secretaria de Estado da Ciência, Tecnologia e Ensino Superior, SETI, Fundo Paraná, and the 2006 Santander Banespa Science and Innovation Prize.

REFERENCES

- Armstrong KF and Ball SL (2005). DNA barcodes for biosecurity: invasive species identification. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1813-1823.
- Ball SL, Hebert PDN, Burian SK and Webb JM (2005). Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *J. N. Am. Benthol. Soc.* 24: 508-524.
- Bensasson D, Zhang D, Hartl DL and Hewitt GM (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* 16: 314-321.
- Boore JL and Brown WM (1998). Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* 8: 668-674.
- Börner C (1925). Lepidoptera, Schmetterlinge. In: Fauna aus Deutschland, 3, verbesserte Auflage (Brohner P, ed.). Quelle and Meyer, Leipzig, 358-387.
- Bravo JP, Felipes J, Zanatta DB, Silva JLC, et al. (2008). Sequence and analysis of the mitochondrial DNA control region in the sugarcane borer *Diatraea saccharalis* (Lepidoptera: Crambidae). *Braz. Arch. Biol. Technol.* 51: 471-477.
- CBOL (2008). Consortium for the Barcodes Life. Available at <http://barcoding.si.edu/>. Accessed January 15, 2008.
- Cywinska A, Hunter FF and Hebert PD (2006). Identifying Canadian mosquito species through DNA barcodes. *Med. Vet. Entomol.* 20: 413-424.
- DeSalle R (2006). Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. *Conserv. Biol.* 20: 1545-1547.
- Fletcher DS and Nye IWB (1984). The Generic Names of the Moths of the World. Vol. 5. Pyraloidea. British Museum (Natural History), London.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, et al. (2006a). DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U. S. A.* 103: 968-971.
- Hajibabaei M, Smith MS, Janzen FH, Rodriguez JJ, et al. (2006b). A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* 6: 959-964.
- Hebert PD, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313-321.
- Hebert PD, Stoeckle MY, Zemlak TS and Francis CM (2004). Identification of birds through DNA barcodes. *PLoS Biol.* 2: e312.
- Holloway JD, Kibby G and Peggie D (2001). The Families of Malesian Moths and Butterflies. E.J. Brill, Leiden.
- Hurst GD and Jiggins FM (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. Biol. Sci.* 272: 1525-1534.
- Kiritani K (1998). Exotic insects in Japan. *Entomol. Sci.* 1: 291-298.
- Krzywinski J and Besansky NJ (2003). Molecular systematics of Anopheles: from subgenera to subpopulations. *Annu. Rev. Entomol.* 48: 111-139.
- Lange CL, Scott KD, Graham GC, Sallam MN, et al. (2004). Sugarcane moth borers (Lepidoptera: Noctuidae and Pyraloidea): phylogenetics constructed using COII and 16S mitochondrial partial gene sequences. *Bull. Entomol. Res.* 94: 457-464.
- Miller SE (2007). DNA barcoding and the Renaissance of taxonomy. *Proc. Natl. Acad. Sci. U. S. A.* 104: 4775-4776.
- Min XJ and Hickey DA (2007). Assessing the effect of varying sequence length on DNA barcoding of fungi. *Mol. Ecol. Notes* 7: 365-373.
- Minet J (1985). Etude morphologique et phylogénétique des organes tympaniques des Pyraloidea. 2 - Pyralidae; Crambidae, première partie (Lepidoptera, Glossata) = Morphological and phylogenetical study of the tympanic organs of the Pyraloidea. 2. Pyralidae, Crambidae, first part (Lepidoptera, Glossata). *Ann. Soc. Entomol. Fr.* 21: 69-86.
- Nelson LA, Wallman JF and Dowton M (2007). Using COI barcodes to identify forensically and medically important blowflies. *Med. Vet. Entomol.* 21: 44-52.

- Rach J, DeSalle R, Sarkar IN, Schierwater B, et al. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proc. Biol. Sci.* 275: 237-247.
- Ratnasingham S and Hebert PDN (2007). BOLD: the barcode of life data system (www.barcodinglife.org). *Mol. Ecol. Notes* 7: 355-367.
- Remigio EA and Hebert PDN (2003). Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. *Mol. Phylogenet. Evol.* 29: 641-647.
- Rubinoff D, Cameron S and Will K (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *J. Hered.* 97: 581-594.
- Sato H, Une Y, Kawakami S, Saito E, et al. (2005). Fatal *Baylisascaris* larva migrans in a colony of Japanese macaques kept by a safari-style zoo in Japan. *J. Parasitol.* 91: 716-719.
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, et al. (2005). Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1805-1811.
- Shaffer M, Nielsen ES and Horak M (1996). Pyraloidea. In: Checklist of the Lepidoptera of Australia. Monographs on Australian Lepidoptera (Nielsen ES, Edwards ED and Rangsi TV, eds.). Vol. 4. CSIRO, Collingwood, 164-199.
- Smith PT, Wood DM, Janzen DH, Hallwachs W, et al. (2007). DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalist. *PNAS* 104: 4967-4972.
- The Barcode of Life Data Systems (BOLD) (2008). Available at <http://www.barcodinglife.org/views/login.php>. Accessed January 15, 2008.
- Ward RD, Zemlak TS, Innes BH, Last PR, et al. (2005). DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1847-1857.
- Wolstenholme DR (1992). Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* 141: 173-216.
- Zhang DX and Hewitt G (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* 25: 99-120.