Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 43 (2007) 530–542
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

A comprehensive DNA sequence library is essential for identification with DNA barcodes

Torbjørn Ekrem [a,*], Endre Willassen [b], Elisabeth Stur [a]

[a] *Section of Natural History, Museum of Natural History and Archaeology, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[b] *Natural History Collections, Bergen Museum, University of Bergen, Muséplass 3, NO-5007 Bergen, Norway*

Received 18 April 2006; revised 24 October 2006; accepted 12 November 2006
Available online 5 December 2006

Abstract

In this study we examine the possibility of utilising partial cox1 gene sequences as barcodes to identify non-biting midges (Diptera: Chironomidae). We analysed DNA from 97 specimens of 47 species in the genera *Cladotanytarsus*, *Micropsectra*, *Parapsectra*, *Paratanytarsus*, *Rheotanytarsus*, *Tanytarsus* and *Virgatanytarsus* with a main focus on *Micropsectra*, *Parapsectra* and *Paratanytarsus*. Our findings show that (1) cox1 is easily amplified from extracts from different life stages with the standard barcoding primers. (2) Although K2P-distances between con-specific sequences varied up to 4.9%, con-specifics clustered together with 91–100% bootstrap support in maximum parsimony analysis. This indicates that barcodes may be excellent tools to identify species that are already in a cox1 library. (3) Both neighbour joining and maximum parsimony failed to reconstruct monophyletic genera. Thus, if a well-matching cox1 sequence is not already available in the library, the prospects of approximately identifying an unknown taxon, even to the correct genus of subtribe Tanytarsina, are not good.

Keywords: DNA-barcoding; DNA library; Chironomidae; *Micropsectra*; *Paratanytarsus*; *Parapsectra*; Cox1; CO1

1. Introduction

The immature stages of the dipteran family Chironomidae are commonly the most diverse and abundant macroinvertebrates in freshwater ecosystems. Many species have specific habitat requirements and the species compositions in diverse fresh waters are often read as pointers to more or less distinct states of environmental gradients. Therefore, chironomids are frequently used by freshwater biologists to assess and monitor environmental conditions (e.g. Brodersen and Lindegaard, 1999; Verneaux and Verneaux, 2002; Aagaard et al., 2004) and to infer past environments from the usually abundant and species rich fossil larval head capsules in lake sediments (e.g. Velle et al., 2005). Unfortunately, the larvae and females of closely related species are usually difficult to distinguish by means of morphology, and species identification frequently depends on association of these life stages with identified pupal exuviae or adult males which tend to posses more species specific characteristics. Moreover, the larvae and females of numerous species, even in the relatively well-documented European fauna, remain unknown to science. Hence, environmental assessments and bio-monitoring of freshwater habitats presumably would have much to gain if the larvae and other life history stages could be more readily identified to species.

Relatively few authors have as yet reported the use of genetic markers as identification tools in studies of Chironomidae. Different molecular techniques and target genes have been used in those few studies (Asari et al., 2004; Carew et al., 2003, 2005; Ekrem and Stur, in press; Sharley et al., 2004; Willassen, 2005) and some of the results may be of limited interest as identification tools for practising

* Corresponding author.
E-mail addresses: Torbjorn.Ekrem@vm.ntnu.no (T. Ekrem), Endre.Willassen@zmb.uib.no (E. Willassen), Elisabeth.Stur@vm.ntnu.no (E. Stur).

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.11.021

freshwater ecologists, who probably prefer methods that are directly comparable and universally applicable for all Chironomidae taxa.

The Barcoding of Life initiative (Hebert et al., 2003a,b) has envisioned a standardized method to alleviate difficult species identifications by focusing sequencing efforts on one target gene, cytochrome *c* oxidase subunit 1 (cox1). Cox1 sequence clustering by neighbour joining (NJ) has been suggested as an effective and suitable way to recognise and identify animal species (Hebert et al., 2003a,b; Hebert et al., 2004a) and to discover cryptic taxa (Hebert et al., 2004b). This approach has been argued to be too imprecise for reliable species diagnoses in some cases, and character based identification systems have therefore been proposed as the preferred way to proceed with the Barcoding of Life Project (DeSalle et al., 2005).

Complete knowledge of the life stages of an organism is essential for a good understanding of its ecology, taxonomy, phylogeny and evolution. An acknowledged advantage of DNA-barcoding is the possibility to easily associate different life stages of the same species (Blaxter, 2004; Stoeckle, 2003). This is particularly valuable when taxa are difficult to rear in the laboratory, and several studies have recognised the benefit of short DNA sequences in associating immature stages with adult counterparts (e.g. Hebert et al., 2004b; Barrett and Hebert, 2005; Miller et al., 2005; Paquin and Hedin, 2004; Thomas et al., 2005; Vences et al., 2005a,b). Many chironomid species are both difficult to rear in the laboratory and unknown as larvae. Only one study has so far demonstrated the usefulness of DNA-barcodes to associate chironomid life stages in practice (Carew et al., 2005).

The goals of the present study were to (1) test the success of cox1 amplification in chironomids with the suggested general PCR primers for DNA-barcoding (LCO1490 and HCO2198) and (2) investigate whether partial cox1 gene sequences can be used to associate life stages and identify species of non-biting midges. We chose to focus on species of the Tanytarsini genera *Micropsectra*, *Parapsectra* and *Paratanytarsus* because the species comprised by these genera are well studied and presumably morphologically relatively distinct. Fresh material of numerous species was available from some of our latest field trips in Europe.

2. Materials and methods

2.1. Taxon sampling and identification

The taxa included in this study were selected with the aim of getting representatives of as many of the known species in *Micropsectra*, *Parapsectra* and *Paratanytarsus* as possible. Field work was mainly conducted in Europe, but some material from other geographical regions was made available to us by colleagues (Table 1). Most of the specimens sampled were adult males ($n = 81$, 85.3%), nine of which were reared from immature stages and thus have associated larval and/or pupal exuviae (Table 1). Six specimens were adult females that all but one could be identified to species by associated pupal skins. Two of the sampled specimens were pupae and six were larvae of which four only could be identified to genus level due to incomplete knowledge of larval taxonomy (*Micropsectra* sp. B, *Micropsectra* sp. C, *Paratanytarsus* sp., *Rheotanytarsus* sp.). Some of the species sampled in this study also could not be named since they were previously unknown to science and are pending formal descriptions (*Micropsectra* sp. A, *Parapsectra* sp. A, *Parapsectra* sp. B, *Tanytarsus* sp. B). More than one life stage was sequenced from six of the included species (Table 1). When possible, we included three specimens of each species, preferably from different locations. The species were identified by their original descriptions and by recent revisions (Stur and Ekrem, 2006; Säwedal, 1976; Reiss and Säwedal, 1981). We also examined types and other reference material in the Natural History Collections in Bergen, Norway and Zoologische Staatssammlung München, Germany. Voucher specimens are deposited in the Natural History Collections, Bergen Museum, University of Bergen, Norway and in the Museum of Natural History and Archaeology, Norwegian University of Science and Technology in Trondheim, Norway.

2.1.1. Extraction, amplification, sequencing and alignment

DNA extraction largely followed the standard protocol for the Qiagen DNeasy tissue extraction kit. We used only 120–170 µl elution buffer (depending on specimen size) to yield an appropriate concentration of DNA in the DNA template solutions. Each PCR was made with addition of 2 µl DNA template, 2.5µl 10 × PCR buffer (Qiagen, with ~15 mM $MgCl_2$), 2 µl of dNTPs in 10 µM concentration, 1 µl of each of the suggested standard barcode primers (Folmer et al., 1994) LCO1490 (5′-GGTCAACA AATCATAAAGATATTGG-3′) and HCO2198 (5′-TA AACTTCAGGGTGACCAAAAAATCA-3′) in 10 µM concentration, 1 U of Qiagen HotStar *Taq* DNA polymerase, and distilled water for a total reaction volume of 25 µl. The PCR had 5 cycles of 30 s annealing at 45 °C and 35 cycles of 30 s annealing at 51 °C in a typical step-up procedure on PTC-100 and PTC-200 PCR machines from MJ Research. The PCR products were purified using QIAquick PCR purification kit (Qiagen). Purified products were sequenced in both directions using BigDye (Perkin-Elmer) termination reactions and analysed on ABI377 or ABI Prism 3100 genetic analysers. Sequences were assembled and edited using Sequencher 3.1.1 (Gibbs and Cockerill, 1995) or BioEdit 7.0.5.2 (Hall, 1999), and aligned in BioEdit. Alignment of the nucleotide sequences was unproblematic since indels were absent and conceptual translation with the invertebrate mitochondrial code returned uninterrupted amino acid sequences that were identified as cox1 fragments with *blastp* search in GenBank. After trimming of uncertain bases at both ends, the aligned sequences were 654 bp long. An overview of species sequenced and their respective GenBank accession numbers is given in Table 1.