

MtDNA COI barcodes reveal cryptic diversity in the *Baetis vernus* group (Ephemeroptera, Baetidae)

Gunilla Ståhls^{a,*}, Eino Savolainen^b

^a Finnish Museum of Natural History, University of Helsinki, P.O. Box 17, FI-00014 Helsinki, Finland

^b Kuopio Natural History Museum, Myhkyrinkatu 22, FI-70100 Kuopio, Finland

Received 6 March 2007; revised 20 August 2007; accepted 7 September 2007

Available online 18 September 2007

Abstract

Partial mitochondrial COI sequences (barcoding fragment) were explored for the understanding of the species boundaries of *Baetis vernus* group taxa (Ephemeroptera, Baetidae) in northern Europe. We sampled all species of this group occurring in Finland, but focused on taxa for which morphological and taxonomical confusion have been most apparent. The sequence matrix comprised 627 nucleotides for 96 specimens, and was analysed using parsimony. Results provided strong evidence that *Baetis macani* Kimmins and *B. vernus* Curtis comprise morphologically cryptic but molecularly distinct taxa, as intraspecific uncorrected divergences within haplogroups ranged between 0.3% and 1.4% and interspecific divergences were from 13.1% to 16.5%. These interesting findings prompt for further taxonomic studies of *B. vernus* taxa using more extensive specimen sampling from the known distributional areas in the Palaearctic/Holarctic region for better understanding of haplotype distributions. We stress the importance of integration of morphological and molecular data, and the necessity to employ additional nuclear DNA sequence data.

© 2007 Elsevier Inc. All rights reserved.

Keywords: *Baetis vernus* group; Ephemeroptera; COI; Barcoding

1. Introduction

From Finland 19 species of mayflies of the family Baetidae are known, and 11 of these belong to the genus *Baetis* (Engblom, 1996). The genus *Baetis* Leach, 1815 (Ephemeroptera, Baetidae) is geographically widespread (Jacob, 2003), and many species are abundant, particularly in northernmost Europe. *Baetis* nymphs, particularly those of *B. rhodani* Pictet, are routinely used in aquatic biomonitoring (e.g. Fialkowski et al., 2003; Williams et al., 2006 and references therein). The *Baetis vernus* group was defined by Müller-Liebenau (1969) and comprises six species in northern Europe, *Baetis bundyae* auctt. nec Lehmkuhl, 1973 (*B. bundyae* is so far not reported from Finland), *B. liebenauae* Keffermüller, 1974, *B. macani* Kimmins, 1957, *B. subalpinus* Bengtsson, 1917, *B. tracheatus*

Keffermüller and Machel, 1967 and *B. vernus* Curtis, 1834. The morphological characters of *Baetis vernus* group nymphs are subject to extensive, evidently environmentally induced morphological variation (Bauernfeind and Hympeesch, 2001) which makes them notoriously difficult to separate even using the authoritative key to northern European taxa by Engblom (1996) (Savolainen et al., 2007). Other determination keys are based on central European material (Müller-Liebenau, 1969) and therefore the usefulness for northern European material is limited. In particular, the nymphs of *B. vernus* and *B. subalpinus* are difficult to separate.

The author Savolainen has studied the distribution and the life history of the *Baetis vernus* group taxa in Finland for over 30 years. The life history of these taxa is recently discussed by Savolainen et al. (2007). Savolainen and Saaristo (1981) pointed out that the taxon *Baetis macani* Kimmins has two forms of nymphs in Finland, one form occurring in lotic waters with narrow gills with invisible tracheae, and another form living in lentic waters and with

* Corresponding author. Fax: +358 9 19144430.

E-mail addresses: Gunilla.stahls@helsinki.fi (G. Ståhls), savolainen@kuopio.fi (E. Savolainen).

wider gills with conspicuous tracheae resembling *B. tracheatus* Keffermüller and Machel. Ingrid Müller-Liebenau, an authority of *Baetis* taxonomy, studied the morphs but regarded both forms as belonging to *B. macani*. *B. macani* was originally described from a pond 800 m ASL from Mount Saana in Finnish Lapland (Kimmins, 1957; Macan, 1957). Outside of alpine ponds it is found in flowing water, particularly in small brooks (Savolainen et al., 2007). *B. tracheatus* was described from River Odra in western Poland (Keffermüller and Machel, 1967) and was later found from an eutrophic *Stratiotes* river in Finnish Lapland (Savolainen and Saura, 1996).

Engblom (1996) recognized the two forms of Savolainen and Saaristo (1980) and called the form with narrow gills and invisible tracheae living in lotic waters *B. bundyae* Lehmkuhl, 1973, a taxon described from tundra ponds in arctic Canada. It has also been found in flowing water elsewhere in arctic Canada (e.g. Cobb and Flannagan, 1980; Harper and Harper, 1997) and the taxon has an extensive distribution extending south along the Rocky Mountains. Engblom (1996) gave small ponds above the tree line as the typical habitat of *B. bundyae* in Europe. Morihara and McCafferty (1979) stated that the morphological characters of *B. bundyae* Lehmkuhl overlapped with the ones of *B. macani* Kimmins and consequently placed it as a subspecies of *B. macani* Kimmins.

Engblom (1996) called the form with broader gills living in lentic waters *B. macani* Kimmins. Recent studies have shown that *B. bundyae sensu Engblom (1996)* is conspecific with *B. macani* Kimmins, and *B. macani sensu Engblom (1996)* is therefore another taxon, possibly a new species (Savolainen et al., 2007), and here referred to as sp. nov. *B. macani* occurs in northern Europe in stony alpine ponds (Kimmins, 1957; Macan, 1957), and in flowing water, particularly in small brooks (Savolainen et al., 2007).

Savolainen et al. (2007) measured the reproductive isolation between taxa of *Baetis vernus* group from northernmost Europe using allele frequencies of enzyme loci. The results indicated that the *B. macani* nymph form with narrow gills and invisible tracheae living in lotic waters was reproductively isolated from the form with broader gills and conspicuous tracheae living in lentic waters, but the power of the method did not suffice to show reproductive isolation of the other *B. vernus* group species. Savolainen et al. (2007) called for the use of DNA characters (both mitochondrial and nuclear sequences) to potentially enhance the understanding of species boundaries within this important group of mayflies.

We test here the informativeness of mtDNA COI sequences (“COI barcodes”) for the understanding of the species boundaries of *Baetis vernus* group taxa, particularly for the taxa for which morphological and taxonomical confusion has been most apparent. We sampled multiple specimens from multiple populations of *B. vernus* group taxa occurring in Finland, and included specimens of *B. bundyae* from the type locality (Rankin Inlet, Canada).

2. Materials and methods

2.1. Specimen collections

Mayfly nymphs of the *Baetis vernus* group were collected by E. Savolainen and J. Viiri from 18 localities, mainly in northern Finland (Fig. 1, Table 1) and stored in individual tubes in 90–100% ethanol. The morphological distinction of *B. subalpinus* and *B. vernus* nymph samples followed the characters given by Engblom (1996), *B. subalpinus* with equally broad first and second mandibular teeth, while *B. vernus* is distinguished by having the first mandibular tooth twice as broad as the second. The specimen identification followed this character, but, during the sampling it was noticed that the *B. vernus* samples showed variability of the width of the first mandibular tooth, it being 1.5–3.5 times as broad as the second one.

2.2. Laboratory methods

Shape and structure of mandibles, abdominal gills and abdominal tergites (and the membranous area between

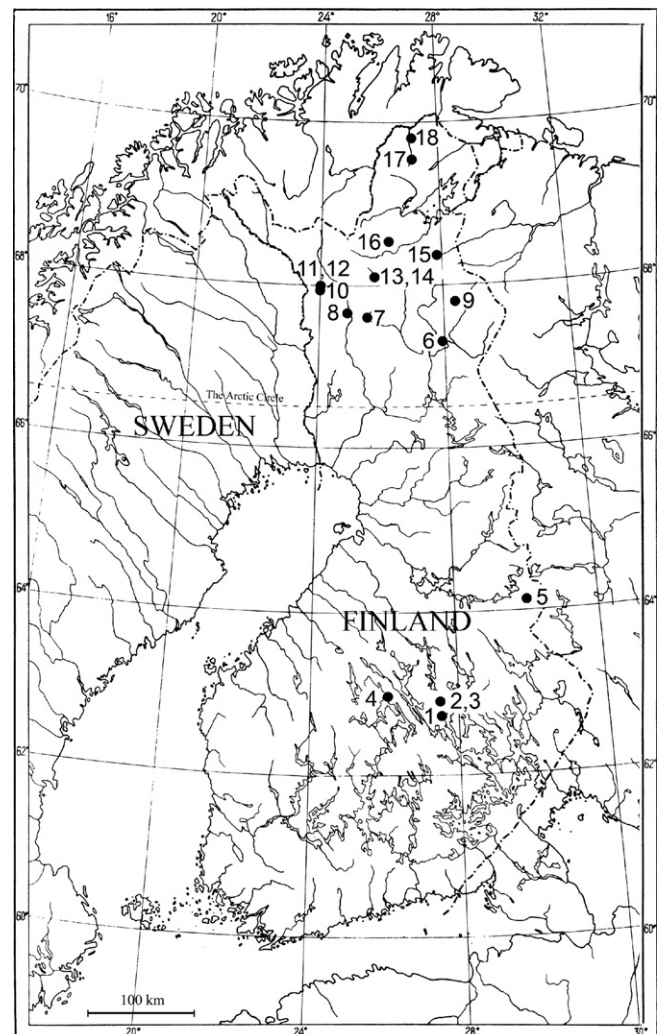


Fig. 1. Map of sampling sites.

Table 1
List of sampling sites

1. Kuopio, Humalajoki	N62°41'36", E27°34'16"
2. Kuopio, Pölkky puro	N62°52'31", E27°37'18"
3. Kuopio, Kallavesi	N62°53'47", E27°36'56"
4. Vesanto, Keitele	N62°55'47", E26°04'45"
5. Kuhmo, Kytöpuro	N64°06'3", E30°5'32"
6. Savukoski, Kuollutoja	N67°18'20", E28°3'47"
7. Kittilä, Jeesiöjoki	N67°39'24", E25°32'43"
8. Kittilä, Myllyjoki	N67°48'54", E24°47'55"
9. Savukoski, Paskalomaaja	N67°52'18", E28°10'50"
10. Muonio, Särkijärvi	N67°54'28", E23°55'36"
11. Muonio, Törmäslommlol	N67°56'59", E24°0'13"
12. Muonio, Kutuniva	N67°57'1", E24°0'18"
13. Kittilä, nameless brook	N68°01'54", E25°43'12"
14. Kittilä, Jalkajoki	N68°5'57", E25°38'58"
15. Inari, Kaunispäänoja	N68°26'27", E27°24'50"
16. Inari, Karvajoki	N68°28'55", E26°4'22"
17. Utsjoki, Molkijoki	N69°30'0", E27°13'38"
18. Utsjoki, Kidesjoki	N69°47'49", E27°5'42"

them) are important in morphological species identification of the nymphs, therefore we used the thorax and legs for DNA extraction. DNA was extracted from these parts of single individuals of the ethanol preserved specimens using the NucleoSpin Tissue DNA Extraction kit (Machery-Nagel, Düren, Germany) following manufacturer's protocols and resuspended in 50 µl ultra-pure water. Remnants of the specimens were conserved for the purpose of further morphological studies and are preserved as DNA voucher specimens, deposited at Finnish Museum of Natural History (MZH).

We used the primer pair LCO1490 (5'-GGTCAA CAAATCATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTCAGGGTGACCAAAAAATCA-3') (Folmer et al., 1994) to amplify the 'Folmer fragment' of the 5'-end of the mitochondrial gene cytochrome *c* oxidase subunit I (COI). PCR amplifications were carried out in 25 µl reactions containing 1–3 µl DNA extract, 1 µl of each primer (at 10 pmol/µl), 0.25 µl of *Taq* DNA polymerase (5 U/µl), 2–3 µl 2.5 mM MgCl₂, 2.5 µl 10× Buffer II (Applied Biosystems, Foster City, CA, USA) and 4 µl 200 mM dNTP (GeneAmp) and ultra-pure water. Thermocycler conditions were initial denaturing at 95 °C 2 min, 29 cycles of 30 s denaturing at 94 °C, 30 s annealing at 49 °C, 2 min extension at 72 °C, followed by a final extension of 8 min at 72 °C.

PCR products were purified using the GFX PCR Purification Kit (Amersham Biotech, Little Chalfont, UK). Amplified PCR was checked for size and products (bands) running 4 µl on a 1% agarose gel and then sequenced (with the PCR primers) in both directions using the Big Dye Terminator Cycle Sequencing Kit vs. 1.1 (Applied Biosystems) at one-fourth of the recommended volumes on an ABI PRISM 377 (Applied Biosystems) sequencer. The sequences were inspected and edited for base-calling errors and assembled using Sequence Navigator™ (version 1.01, Applied Biosystems). All sequences were submitted to

EMBL Nucleotide Sequence Database with Accession Nos. AM494594–AM494688.

2.3. Parsimony analysis

We used the program NONA v2.0 (Goloboff, 1993) for the parsimony analyses (command line: hold 100,000; mult*1000; hold/200) and for calculating the Bremer support (branch support) values. All characters were equally weighted, and gaps were treated as missing data. *Baetis (Nigrobaetis) niger* (Linnaeus, 1761) was used as outgroup.

3. Results

3.1. Sequences

We obtained a molecular dataset of 627 nucleotides for 96 specimens of the *B. vernus* group, and four specimens of the outgroup *Baetis (Nigrobaetis) niger*. Mean base frequencies were A = 22.5%, C = 18.2%, G = 22.4% and T = 37.0%. Uncorrected pairwise divergences between ingroup taxa ranged between 4.61% and 17.1%, and between outgroup and ingroup taxa from 17.7–21.2%.

We also tested PCR amplifications of parts of the nuclear ribosomal cluster with universal primers commonly used for various insect orders. Primers for the ITS2, D2-3 domain of 28S rRNA gene and a fragment of the 18S rRNA gene were tested. Amplification was not successful for the ITS2 region. The highly conserved 18S fragment amplified well, but showed nucleotide variation only between the outgroup and the ingroup taxa, and was therefore not informative for the present study. The D2 region of the 28S rRNA gene amplified poorly, and we failed to produce good quality sequences with the primers that were tested.

3.2. Parsimony analysis

No insertions or deletions occurred in the COI dataset so alignment was unambiguous. Of the obtained 627 nucleotides, 208 sites were parsimony-informative. Parsimony analysis produced 1620 equally parsimonious trees of 492 steps in length, with a consistency index (CI) of 0.63 and a retention index (RI) of 0.96. The strict consensus is shown in Fig. 2.

All *B. liebenauae* specimens from the sampled population showed identical sequences and clustered together. The sister group was a clade comprising part of the *B. vernus* specimens (Kuopio, Kuhmo and Kittilä, Jeesiöjoki populations, sample sites 2, 5 and 7, Fig. 1). The intraspecific variation of *B. vernus* clade 1 is low, at most four nucleotide changes.

All included specimens sampled from five populations (sample sites no 3, 4, 7, 10, 11) of the sp. nov. (the lotic form of *B. macani*), were resolved in one clade. Again, intraspecific variability was low. The uncorrected pairwise

4. Discussion

4.1. The taxonomic status of *B. vernus* group taxa

The enzyme electrophoresis study of Savolainen et al. (2007) showed that the lentic form of *Baetis macani* with broader gills and conspicuous tracheae and the lotic form *B. macani* with narrower gills and invisible tracheae were reproductively isolated from each other, and in the NJ-phenogram consistently fell within separate clusters. Our results support their conclusion of the lentic form being a separate species, with average uncorrected pairwise divergence of 14.3% between these taxa (Fig. 2).

Savolainen et al. (2007) found only low genetic differentiation between *B. macani/subalpinus/vernus*, and hypothesized that this could be a consequence of gene flow across species boundaries (introgression). We found clear differences between these taxa, as measured with uncorrected pairwise divergences ranging between 14.0% and 16.9%. The new taxon (taxa) will be described elsewhere (Savolainen, in preparation) Ball et al. (2005) recorded K2P distances between four *Baetis* species belonging to different subgenera, and found a range between 8.5% and 24.6% (mean 20.2%). Williams et al. (2006) in their study on COI diversity of the geographically wide-spread *Baetis rhodani* Pictet, found that this taxon comprised seven distinct COI haplogroups differing 8–19%. Williams et al. stated that these haplogroups represent several evolutionary lineages whose current species status is undefined, but they proposed them as distinct taxonomic units from *B. rhodani* sensu stricto and *B. vernus*. Noticeably, the uncorrected pairwise divergence of three *B. vernus* specimens used by Williams et al. (from UK, Wales) is closest to our *B. liebenauae* (3.67%), and differs from our *B. vernus* groups by 6.1–12.4%.

Based on current knowledge of species level COI divergences in insects, the levels reported by Williams et al. for *B. rhodani* haplogroups strongly suggests separate, morphologically cryptic, taxa. Their interspecific divergences are consistent with the ranges we found for the two *B. vernus* clades and the *Baetis* sp. nov in the present study, and our hypotheses are congruent with those of Williams et al. (2006). Interestingly, all above reported pairwise uncorrected sequence divergence values of COI for both closely or more distantly related species pairs of mayflies belong to the higher extreme, and surpassing maxima typically reported in other insect species (e.g. Roe and Sperling, 2007).

4.2. Taxonomic utility of COI barcodes

Ball et al. (2005) tested the ability of COI profiles to generate correct species identifications for a representative set of North American mayflies, and reported a highly successful result of the test specimens being identified correctly to species by comparisons the COI profiles. Their taxon sampling did not, however, target cases of difficult or confused

morphological taxonomy, the cases where molecular information frequently raises as many questions as it resolves. Other aquatic taxa such as daphniid and rotifer species, that exhibit marked phenotypic plasticity, have been distinguished reliably using COI (Derry et al., 2003; Adamowicz et al., 2004). Problems using DNA barcodes in insects and other invertebrates have been frequently observed, including mitochondrial introgression between taxa, recent speciation followed by incomplete lineage sorting or interbreeding that obscures species identification and/or delimitations (e.g. Croucher et al., 2004; Bachtrog et al., 2006; Kaila and Ståhls, 2006).

Smith et al. (2006) test of the barcoding concept in Diptera applied COI data to a diverse assemblage of *Belvosia* Robineau-Desvoidy, 1830 parasitic flies (Tachinidae). Their conclusion was that morphological concepts are largely supported by molecular data, but that morphology alone does not discover the true diversity hidden within some species groups. This finding is very clearly paralleled by Williams et al. (2006) and by our results, as morphologically cryptic lineages show very distinct molecular differences.

We observed some discrepancies between morphological species assignment and the obtained COI sequence. This was observed in one *B. macani* specimen (ES85) and one *B. liebenauae* specimen (ES154) that were resolved into the second *vernus*-clade. The identity of these specimens was rechecked, and did not change in any of the cases. Thus we suggest mitochondrial introgression in these sampled individuals of sympatric populations of these closely related taxa (see e.g. Bachtrog et al., 2006).

As taxonomic expertise in mayflies is limited, and morphological identification often is difficult (e.g. damaged specimens or fragments of specimens) we envision that a DNA-based identification system based on COI profiles, correlated with sequence data from other loci and informative morphological characters, could provide an important tool for species identification in aquatic biomonitoring of mayflies. The establishment of such a system should be based on data from multi-locus data of both mitochondrial and nuclear gene regions (for avoiding the potential pitfalls of introgression or incomplete lineage sorting). Geographically widely sampled specimens are vital to assess the potentially extensive COI haplotype diversity, as this was observed by Williams et al. (2006) and in many taxa of the present study. And most importantly, a DNA-based system can not replace the involvement and expertise of taxonomists to identify specimens, and to deal with the formal taxonomical descriptions of species new to science.

Our results highlight the utility of mtDNA sequence characters for revealing the hidden diversity and thus potentially also for the understanding of species boundaries of this group of mayflies, but we also acknowledge the clear limitations of using a single molecule. A forthcoming study aims at sampling specimens over a broad geographic range in the Palaearctic (and in some cases Hol-

arctic) area for molecular studies, and at adding DNA characters from an independent nuclear gene region.

Acknowledgments

We thank Johanna Viiri (Kuopio) for help with material collection and initial laboratory work, Anssi Saura (Umeå, Sweden) for discussions and information and Jouko Sarvala (Turku) for help for establishing contacts with Canadian researchers, and Steven Burian (USA) and Donna Gibberson (Canada) for sending specimens of *B. bundyae* for molecular work. Kuopion Luonnonystävien Yhdistys and Finnish Ministry of Environment (PUTTE research program and Aquatic Insect Expert group) provided financial support for the laboratory work and collecting trips.

References

- Adamowicz, S.J., Hebert, P.D.N., Marinone, M.C., 2004. Species diversity and endemism in the *Daphnia* of Argentina. *Zoological Journal of the Linnean Society* 140, 171–204.
- Bachtrog, D., Thornton, K., Clark, A., Andolfatto, P., 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution* 60, 292–302.
- Ball, S.L., Hebert, P.D.N., Burian, S.K., Webb, J.M., 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. *Journal of the North American Benthological Society* 24, 508–524.
- Bauernfeind, E., Hymesch, U.H., 2001. Die Eintagsfliegen Zentraleuropas (Insecta: Ephemeroptera): Bestimmung und Ökologie., Verlag des Naturhistorischen Museums Wien, Wien, 239 pp.
- Cobb, D.G., Flannagan, J.F., 1980. The distribution of Ephemeroptera in northern Canada. In: Flannagan, J.F., Marshall, K.E. (Eds.), *Advances in Ephemeroptera Biology*. Plenum Press, New York, pp. 155–166.
- Croucher, P.J.P., Oxford, G.S., Searle, J.B., 2004. Mitochondrial differentiation, introgression of species in the *Tegenaria atrica* group (Araneae: Agelenidae). *Biological Journal of the Linnean Society* 81, 79–89.
- Derry, A.M., Hebert, P.D.N., Prepas, E.E., 2003. Evolution of rotifers in saline and subsaline lakes: a molecular phylogenetic approach. *Limnology and Oceanography* 48, 675–685.
- Engblom, E., 1996. Ephemeroptera (Mayflies). In: Nilsson, A.N. (Ed.), *The Aquatic Insects of North Europe*. Apollo Books, Steenstrup, pp. 13–53.
- Fialkowski, W., Klonowska-Olejnik, M., Smith, B.D., Rainbow, P.S., 2003. Mayfly larvae (*Baetis rhodani* and *B. vernus*) as biomonitors of trace metal pollution in streams of a catchment draining a zinc and lead mining area of Upper Silesia Poland. *Environmental Pollution* 121, 253–267.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3, 294–299.
- Goloboff, P.A. (1993) NONA. Noname (a bastard son of Pee-wee), version 2.0 (32 bit version). Available from: <http://www.cladistics.com/aboutNona.htm> (accessed 6 February 2006).
- Harper, P., Harper, F.F., 1997. Mayflies (Ephemeroptera) of the Yukon. In: Danks, H.W., Downes, J.A. (Eds.), *Insects of Yukon. Biological survey of Canada (Terrestrial Arthropods)*, Canadian Museum of Nature, Ottawa, pp. 151–168.
- Jacob, U., 2003. *Baetis* Leach 1815, sensu stricto oder sensu lato Ein Beitrag zum Gattungskonzept auf der Grundlage von Artengruppen mit Bestimmungsschlüsseln.. *Lauterbornia* 47, 59–129.
- Kaila, L., Ståhls, G., 2006. DNA barcodes: evaluating the potential of COI to differentiate closely related species of *Elachista* (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa* 1170, 1–26.
- Keffermüller, M., Machel, M., 1967. *Baetis tracheatus*, sp. n. (*Ephemeroptera, Baetidae*).. *Bad. Fizjograf. Pol. Zach.* 20, 11–14.
- Kimmins, D.E., 1957. A new species of the genus *Baëtis* (Ephemeroptera) from North Finland. *Notulae Entomologicae* 37, 27–29.
- Macan, T.T., 1957. A description of the nymph of *Baëtis macani* Kimmins. *Notulae Entomologicae* 37, 58–60.
- Moriyama, D.K., McCafferty, W.P., 1979. Subspecies of the transatlantic species, *Baetis macani* (Ephemeroptera: Baetidae). *Proceedings of Entomological Society of Washington* 81, 34–37.
- Müller-Liebenau, I., 1969. Revision der europäischen Arten der Gattung *Baetis* Leach, 1815. *Limnologische Schriftenreihe*, 1–214.
- Savolainen, E., Drotz, K.M., Hoffsten, P.-O., Saura, A., 2007. The *Baetis vernus* group (Ephemeroptera, Baetidae) of northernmost Europe: an evidently, diverse but poorly understood group of mayflies. *Entomologica Fennica*. 18, 160–167.
- Savolainen, E., Saaristo, M.I., 1980. *Baetis digitatus* Bengtsson (Ephemeroptera) found in Finland. *Notulae Entomologicae* 60, 195–196.
- Savolainen, E., Saaristo, M., 1981. Distribution of mayflies (Ephemeroptera) in the biological province of Kuusamo (Ks), Finland. *Notulae Entomologicae* 61, 117–124.
- Savolainen, E., Saura, A., 1996. *Baetis tracheatus* Keffermüller and Machel (Ephemeroptera, Baetidae), the first record in Finland; found in an eutrophic Stratiotes-river in Kittilä. *Sahlbergia* 3, 28–29.
- Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America* 103, 3657–3662.
- Williams, H.C., Ormerod, S.J., Bruford, M.W., 2006. Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and Evolution* 40, 370–382.