NTNU – Trondheim
Norwegian University of
Science and Technology

Phylogenetic analysis of the genus *Paratanytarsus* (Diptera: Chironomidae)

Sondre Dahle

Biology

Supervisor: Torbjørn Ekrem, IBI
Co-supervisor: Elisabeth Stur, Vitenskapsmuseet, Seksjon for naturhistorie

Norwegian University of Science and Technology
Department of Biology

Preface

This thesis has been carried out at the Norwegian University of Science and Technology (NTNU), Museum of Natural History and Archaeology, with the supervision of Associate Professor Torbjørn Ekrem and researcher Elisabeth Stur. Part of the material for this study was collected during fieldwork in collaboration with “Finnmarksprosjektet” financed by the Norwegian Taxonomy Initiative.

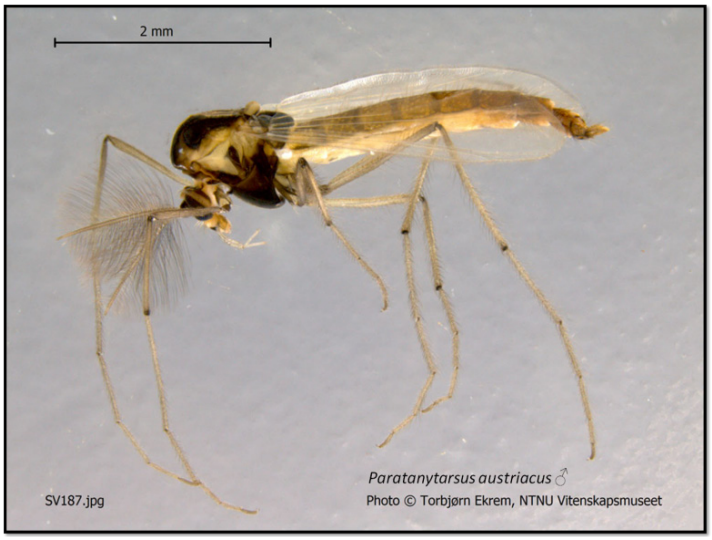
I want to thank my two supervisors, who have helped me through this period with encouragements and excellent supervision. Also, thanks to Rossana Segreto and Renate Bardal for help with the lab work, to Peter Cranston for providing Australian material, Brian Cassel for primer sequences and to Alyssa Anderson and Fabio Laurindo da Silva for helpful comments and corrections on the manuscript.

Enjoy!

Front-page: Portrait of a mature *Paratanytarsus austriacus* female

Sammendrag

Fjærmygg (Chironomidae) er en av de mest suksessrike insektgrupper i akvatiske habitater og finnes i de fleste typer ferskvannsmiljøer. Fjærmyggslekten *Paratanytarsus* er kjent fra alle biogeografiske regioner unntatt tropisk Afrika, 19 arter er registrert i Europa. Tidligere molekylære undersøkelser har antydet at det finnes ubeskrevne arter i slekten, i tillegg har det blitt stilt spørsmål ved monofylien til gruppen. I denne studien blir fire nukleære molekylære markører , CAD1, CAD4, PGD og AATS1 brukt for å kartlegge slektskapet mellom individer bestemt til 16 forskjellige arter, samlet ved lokaliteter i nord Europa, Arktisk Canada og Australia. De fylogenetiske resultatene støtter slektens monofyli, mens intraspesifikk variasjon i flere grupper foreslår tilstedeværelse av ubeskrevne arter. Materiale bestemt til *P. austriacus/hyperboreus* inneholder fire genetisk separate linjer og en ny kryptisk art for Norge. Canadiske *P. dissimilis* og *P. tenuis* grupperer parafyletisk med individer bestemt til samme art fra Europa og kan representere ubeskrevne arter. Den inkluderte Australske arten ender opp mellom taxa fra den nordlige halvkule, men uten nære slektninger. Det spekuleres i om dette kan være et resultat av bipolare migrasjoner i slektens historie.



Abstract

The non-biting midges (Chironomidae) are among of the most successful insects in freshwater systems and often dominate in abundance and species richness. The genus *Paratanytarsus* contains species from all biogeographic regions except tropical Africa, 19 species are known from Europe. Previous molecular work has suggested the presence of undescribed species within some species groups, in addition the monophyly of the genus has been questioned. In this study four nuclear molecular markers, CAD1, CAD4, PGD and AATS1 are utilized in order to reconstruct the evolutionary history of the genus. Samples identified to 16 different species have been collected at locations in Northern Europe, Arctic Canada and Australia. The results of the phylogenetic analysis supports the monophyly of the genus, while considerable intraspesific variation is revealed within several species. Material identified to *P. austriacus/hyperboreus* is found to group into four separated genetic clusters, two of which appear to be undescribed cryptic species based on currently used morphological characters. Canadian *P. dissimilis* and *P. tenuis* ends up paraphyletic with respect to European samples and might represent new Nearctic species. The Australian taxa came out well-embedded in the tree without any close relatives. It is hypothesized that bipolar migrations has occurred in the history of the genus.

Contents

1. Introduction

One of the major challenges for systematists and taxonomists is the recognition and delimitation of species and reconstruction of their phylogenetic relationships (Wiens 2007). This is especially relevant and challenging for insects, which make up most of the described species on Earth. The non-biting midges (Chironomidae) belong to the diverse insect order Diptera and are the most widely distributed and frequently the most abundant group of freshwater insects (Cranston 1995). It is estimated that more than 10,000 species exists worldwide (Cranston 1995); currently, 1,200 are recorded from Europe (Sæther et al. 2010). Due to their high diversity, they are very useful in biomonitoring of freshwaters (Raunio et al. 2007) and in climate reconstruction because of the preservation of sclerotised larval head capsules in lake sediments (Velle et al. 2005). Different species vary in their environmental response and habitat preferences, accordingly it is important to have an established phylogeny to correctly infer species boundaries. This thesis focuses on the phylogeny of one genus, *Paratanytarsus* Thienemann & Bause, 1913 within the family Chironomidae.

The genus *Paratanytarsus* is one of about 30 recognized genera in the tribe Tanytarsini which is one of three tribes in the subfamily Chironominae (Sæther et al. 2004). *Paratanytarsus* was introduced by Thienemann and Bause (Bause 1913) as a genus containing the *lauterborni*-group and the Attersee-*Tanytarsus* and "certainly other subgenera", but without any generic diagnosis.
Paratanytarsus was later included as one of seven subsections in "section Tanytarsus genuinus" by Thienemann (1951) where it contained the genera *Diatanytarsus* Kieffer, *Stylotanytarsus* Kieffer, *Paratanytarsus*, and *Monotanytarsus* Kieffer. This was largely based on characters in the pupal stage and it was shown by Brundin (1947; 1949), Palmén (1960) and Reiss (1968b) that this classification was not tenable for the adult males. The genera *Monotanytarsus* and *Ditanytarsus* were later treated as synonyms of *Paratanytarsus* (Reiss 1974; Reiss et al. 1981). An earlier record of the name *Ditanytarsus,* used to designate a species in the genus *Tanytarsus*, was later discovered resulting in a synonymization with *Tanytarsus* (Ashe 1983). It has been suggested that *Stylotanytarsus* might qualify to be ranked as a genus by Reiss and Säwedal (1981), although this was not supported by Langton et al. (1988) who treated *Stylotanytarsus* as a junior synonym of *Paratanytarsus*.

Morphologically, adults of *Paratanytarsus* resemble males of *Micropsectra* Kieffer, but can be separated by having tibial combs with spurs, whereas in *Micropsectra,* spurs are

almost always absent. The shape of the anal crests is also characteristic; in *Paratanytarsus* they are short and rounded, while in *Micropsectra* they are longer and parallel or convergent (Cranston et al. 1989). The genus can also be separated in the pupal stage by the arrangement of point or spinule patches on tergites III – VI and by the presence of a pearl row on the margin of the wing sheath (Pinder et al. 1986). In the larval stage, the genus forms a morphologically well-defined group recognized by the pecten epipharyngis which consists of 3-5 separate finger-like lobes and by having Lauterborn organs on sessile or very short pedicels (Pinder et al. 1983). These characters has recently been challenged by the discovery of a new species, *P. longistilus* Bolton et al. 2010, that possesses new combinations of characters in all life stages. The adult male lacks crests on the anal point, the pupa does not have a pearl row on the wing sheath, and in the larval stage, the pedicel of the Lauterborn organ is longer than antennal segments 3-5 combined and the pectin epipharyngis is composed of three plates (Bolton et al. 2010). Two newly described species from Brazil, *P. corbii* Trivinho-Strixino 2010 and *P. silentii* Trivinho-Strixino 2010, also deviate slightly for the former established genus characters by lacking anal point crests (Trivinho-Strixino 2010).
Adult females are mostly undescribed, with the notable exception of *P. grimmii* Schneider 1885, a species only known from females due to its parthenogenetic life cycle. This species sometimes occurs as a pest, because of its capacity to breed in water distribution systems (Langton et al. 1988).

Based on pupal morphology Pinder and Reiss (1986) divided *Paratanytarsus* into several species groups which were also found to be sustainable by Cranston et al. (1989) on the basis of adult male morphology.
The phylogenetic placement of *Paratanytarsus* with regard to its neighboring taxa has been evaluated in a few morphologically based studies. Säwedal (1982) found *Paratanytarsus* to be most related to the genera *Micropsectra*, *Parapsectra* Reiss and *Krenopsectra* Reiss respectively. This picture was also supported by Sæther and Roque (2004), although without evaluation of relatedness. Based on the evidence from molecular markers, Ekrem et al. (2010) synonymized *Krenopsectra* and *Parapsectra* with *Micropsectra* and kept *Micropsectra* as the closest described relative of *Paratanytarsus*.

Since *Paratanytarsus* was introduced, many new species have been described from all continents except Antarctica, adding up to about 50 published species worldwide as of 2012. The distribution is almost certainly worldwide (Cranston et al. 1989). Säwedal (1982) mentions the presence of *Paratanytarsus* from the Afrotropical region without giving any records, although according to Pinder and Reiss (1986), no species are known from this region. As most species of *Paratanytarsus* have been recorded in the northern

hemisphere, it has been hypothesized that the genus has its origin in the Holarctic region (Reiss et al. 1981) and has spread to the southern hemisphere through secondary dispersal (Säwedal 1982).
The habitats of *Paratanytarsus* are quit variable and the larvae and pupae have been found in many types of water bodies, including lakes, rivers, ponds, ditches and occasionally brackish water (Cranston et al. 1989) and cool streams (Bolton et al. 2010).

1.2. Phylogenetic work on Chironomidae

There is a long tradition of phylogenetic work on non-biting midges (Ekrem et al. 2010). More than 300 recognizable genera have been described (Ashe et al. 1987) and the relationship between the genera and their adjacent tribes and subfamilies have been treated in many studies based on morphologically (see Cranston et al. 2012 for a summary).
During the last decade, advancements in DNA sequencing technologies, bioinformatics and computational biology has provided large amounts of molecular sequence data, and improvements to the tools used to analyze them. This new set of characters has provided a basis for reexamination of many controversies in phylogeny and has increasingly improved our understanding of insect relationships (Trautwein et al. 2012). There is a growing body of literature describing the use of molecular data to infer chironomid phylogenies e.g. (Makarevich et al. 2000; Guryev et al. 2001; Cranston et al. 2002; Martin et al. 2002; Papoucheva et al. 2003; Ekrem et al. 2004; Stur et al. 2006; Martin et al. 2007; Cranston et al. 2010; Huang et al. 2011; Krosch et al. 2011). The results often support existing relationships based on morphology, although some polyphyletic groups and new relationships have been discovered (Cranston et al. 2010; Ekrem et al. 2010; Krosch et al. 2011; Cranston et al. 2012). Of these molecular studies, a few included one or more *Paratanytarsus* species:
In Cranston et al. (2012) the two *Paratanytarsus sp.* that were included came out as a well-supported monophyletic group based on data from four genes, two ribosomal (rDNA) genes, 18S and 28S, the nuclear protein coding gene fragments of CAD1 and CAD4 and the mitochondrial DNA (mtDNA) of COI.
Ekrem (2004) included four species, *Paratanytarsus grimmii, P. austriacus* Kieffer, 1924*, P. hyperboreus* Brundin, 1949 and *P. natvigi* Goetghebuer, 1933. Using the COII mtDNA marker, the genus came out as monophyletic with third codon position excluded, although there were indications of a paraphyletic *Paratanytarsus* when all codons were included, probably due to saturation of the phylogenetic signal in this marker (Ekrem et al. 2004). Later, Ekrem et al. (2010) utilized CAD1, the ribosomal 16S and mtDNA COI and COII on ten different *Paratanytarsus* species in addition to species from related genera. Here, *Paratanytarsus* was found to be monophyletic only with the use of CAD

and for the concatenated dataset of all four markers, but with low branch support. One species, *Paratanytarsus austriacus* came out as paraphyletic with regard to the most related taxa, indicating presence of a cryptic species.

1.3. Choice of genetic loci

Different genome regions are known to evolve differently. In *Drosophila*, considerable variation in the evolutionary rate among protein-coding genes have been found, depending on protein function, expression level and chromosomal location (Clark et al. 2007). When it comes to mitochondrial genes (mtDNA), they typically have a synonymous substitution rate that is 4.5-9 times higher than the average for nuclear genes (Moriyama et al. 1997). The rapid mutation rate is one of the reasons why mtDNA is has gained popularity in biodiversity studies and for identifying animal species through so called DNA barcoding (Hebert et al. 2003), using a standardized region of the cytochrome c oxidase subunit I gene (COI). However, for phylogenetic purposes, to be used as a marker to track evolutionary patterns, to high mutation rates can lead to multiple substitutions at the same site which, depending on the accuracy of the model of sequence evolution, can go undetected or be wrongly inferred, in turn misleading the placement of branches on the phylogenetic tree (e.g. Philippe et al. 2011). On the other hand, the evolutionary rate for a potential marker should not be too low, especially when considering speciation events that are closely spaced in time. Otherwise, the phylogenetic signal will be weak and result in few characters and short branches that are difficult to resolve (e.g. Saitou et al. 1986).

For Chironomidae, the COI mitochondrial gene has successfully been used for barcoding purposes (Carew et al. 2007; Ekrem et al. 2007), but in phylogenetic analyses, it has been shown to possess too much variation (Ekrem et al. 2010).

Ribosomal genes, like the 16S and 28S, have for a long time been used to infer insect relationships and many studies have demonstrated their usefulness, but they can generate results that are highly dependent on the alignment method and character inclusion (Trautwein et al. 2012). More unambiguous alignments can be archived by using protein-coding nuclear genes that contain easily aligned coding regions, although these genes are often are more difficult to amplify than ribosomal genes and the presence of multiple gene-copies and large introns can cause problems (Danforth et al. 2005). As more and more sequenced insect genomes have become available, it has become easier to develop taxon-specific primers for conserved genes to resolve these problems (Trautwein et al. 2012). One of the most frequently used nuclear protein-coding genes for dipteran phylogeny is the CAD gene, specifically the carbomyl phosphate synthase domain, introduced as a phylogenetic marker by Moulton (2004)

and shown by Ekrem et al. (2010) to possess a strong phylogenetic signal for relationships within genera of Chironomidae.

1.4. Objectives

So far, no in depth molecular phylogenetic analysis focusing on *Paratanytarsus* has been carried out and the validity of the species groups established on the basis of morphological information has yet to be tested with molecular data. The monophyly of the genus has been questioned using molecular data (Ekrem et al. 2010), and needs to be investigated with more markers.

With this background the following questions are raised:

1. Is *Paratanytarsus* a monophyletic group?
2. What is the relationship between the investigated species?
3. Should the apparently cryptic species in the *austriacus* group be treated as true species?
4. Based on phylogeny, what is the most probable biogeographical history of the genus?

To answer this, the phylogenetic utility of the tree nuclear protein coding genes, phosphogluconate dehydrogenase (PGD), alanyl-tRNA synthetase (AATS1) and triose phosphate isomerase (TPI) is explored. These genes have been used to resolve higher level insect (Wiegmann et al. 2009) and Diptera phylogenies (Wiegmann et al. 2011), however their usefulness in lower level chironomid relationships has yet to be tested. Eventually the two most suitable ones will be combined with CAD1 and CAD4, two regions of CAD, to try to reconstruct the evolutionary history of the genus.

2. Materials and methods

2.1. Data collection

2.1.1. Sampling

Exemplars of 15 described species of *Paratanytarsus* and 5 outgroup taxa from the genera *Micropsectra, Parapsectra* and *Tanytarsus* were chosen for molecular analysis. Several projects in the last few years have generated fresh material stored at the museum the NTNU Museum of Natural History and Archaeology in Trondheim. Some of these projects have involved the use of the COI "barcode gene" for species within *Paratanytarsus* (Ekrem et al. 2007; Ekrem et al. 2010; Ekrem et al. 2010; Stur et al. 2011); distance trees produced from this material facilitated the selection of material to be included within this research. The goal when selecting material was to include all the species that were available and with the help of COI sequences, include material from the different genetic clusters within what was identified as the same species. In cases where a species had been sampled at geographically well-separated locations, samples from both locations were included. The identified material mostly consisted of adult males, in addition, a few *Paratanytarsus* females and larvae unidentified to the species level were included, making a total molecular dataset of 46 specimens from the genus *Paratanytarsus*.

Three species of *Micropsectra, M. nana* Meigen, 1818*, M. borealis* Kieffer, 1922 and *M. roseiventris* Kieffer, 1909 were chosen to test the monophyly of *Paratanytarsus* based on Ekrem et al. (2010) which found *M. roseiventris* to be the *Micropsectra* species most closely related to *Paratanytarsus.* Two species of *Tanytarsus* were used to root the tree. The new material for this study was identified using taxonomic reviews and original descriptions (Brundin 1949; Reiss 1968b; Reiss et al. 1971; Reiss et al. 1981; Gilka et al. 2010). All other material was identified by E. Stur (Museum of Natural History and Archaeology, NTNU), T. Ekrem (Museum of Natural History and Archaeology, NTNU), W. Gilka (Department of Invertebrate Zoology, University of Gdansk, Poland) and P. Cranston (Evolution, Ecology and Genetics, Australian National University, Australia). Although the molecular analysis sometimes suggested alternative species identifications, the morphological identifications were retained.

The molecular reference material generated for this study, as well as the majority of the mounted specimens, are deposited in the NTNU Museum of Natural History and Archaeology in Trondheim. The remaining specimens are deposited in Bergen Museum, University of Bergen. Data on the included samples is listed in Appendix 1: Table 1.

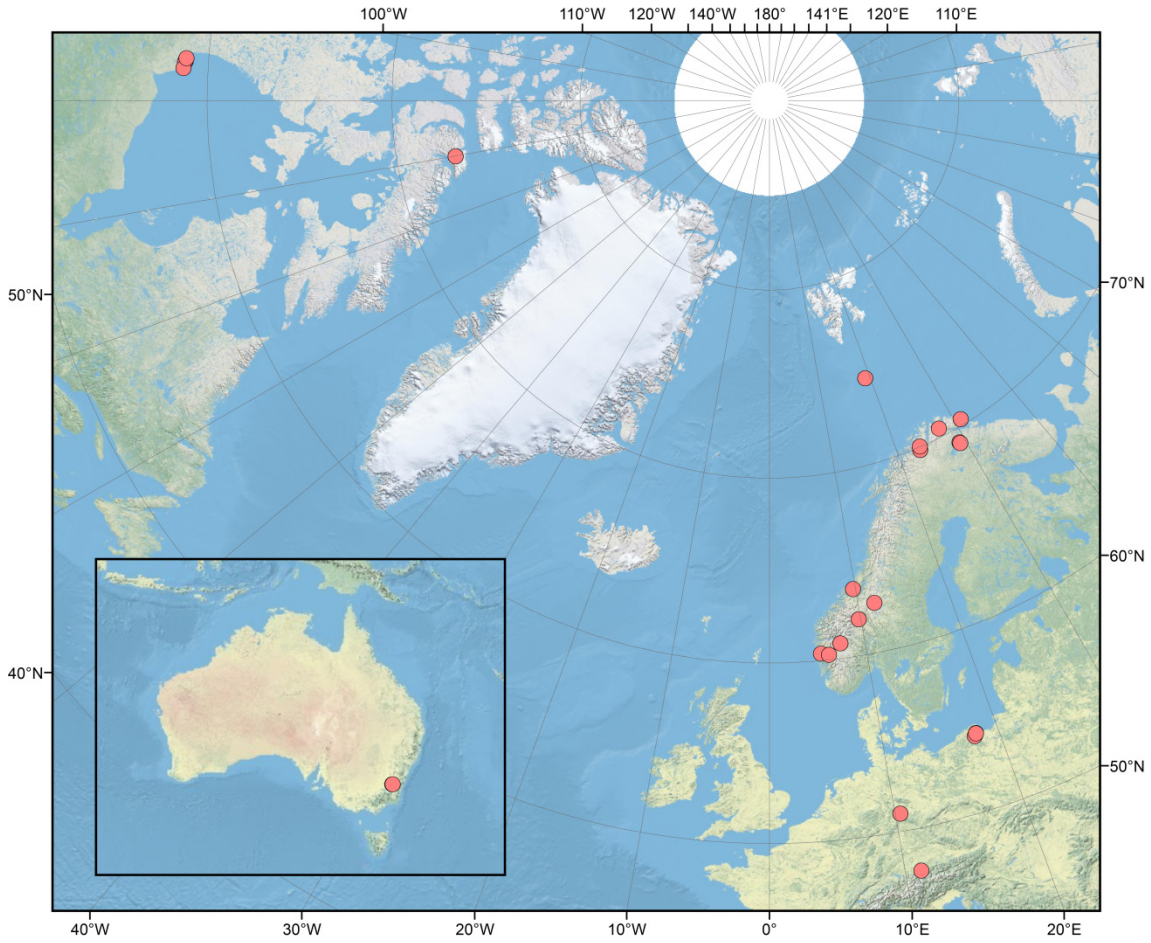


Figure 1, Map showing the collection sites of the *Paratanytarsus* specimens included in this study.

2.1.2.DNA Extraction

Paratanytarsus specimens used for DNA extraction were fixed in ethanol (75-96%) and stored cold (4°C) and dark. The wings, antennae and legs were removed and mounted in Euparal® on microscope slides. DNA was extracted and isolated from the thorax, head and abdomen using Omega BioTek Tissue Extraction Kit D3396-02 or GeneMole DNA Tissue Kit (MG11-102) on a GeneMole® extraction robot (Mole Genetics). When using the Omega BioTek kit, extractions were modified from the standard protocol for animal tissue by using Proteinase K (QIAGEN) instead of OB Protease (20 µL). The lysis was done overnight (≥ 12 h) to ensure complete lysis and to allow for diffusion of DNA out of the exoskeleton. The final elution was reduced to 150 µL due to the small size of the samples.

When using GeneMole DNA Tissue Kit, the standard protocol was followed with the exception that twice the amount of Proteinase K (4 µL) was used with the buffer during lysis; the final elution volume was 100 µL.

For each extract the DNA concentration and purity (A260/280 ratio) was measured using a NanoDrop 2000 (Thermo Scientific) spectrophotometer.

After DNA extraction, the exoskeleton was washed in ethanol (96 %) and mounted on a microscope slide containing the wings, legs and antennae from the same specimen for later morphological analyses.

2.1.3. Amplification

DNA amplifications of five segments (Table 1) was carried out using 2.5 µL 10 x Ex Taq Buffer, 2 µL dNTP Mix, 0.1 µL Ex Taq HS (all TaKaRa Bio INC, Japan) and 1 µL of each 10 µM primer. The amount of template DNA was adjusted according to the DNA concentration and varied between 2-5 µL. Additional 25 mM $MgCl_2$ (0.5 – 1.5 µL) was added for amplification of PGD and AATS1 and for a few samples of CAD. ddH_2O was added to make a total of 25 µL for each reaction.

Amplification cycles were performed on a Biorad C1000 Thermal Cycler and followed a three-step touchdown program with an initial denaturation step of 98 °C for 10 s, then 94 °C for 1 min followed by 5 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 1 min 30 s, 5 cycles of 94 °C for 30 s, 52°C for 30 s, 72 °C for 1 min 30 s and 35 cycles of 94 °C for 30 s, 45 °C for 30 s, 72 °C for 1 min 30 s and one final extension at 72 °C for 3 min. This program was sometimes modified to include 40 cycles in the last step if needed for sufficient amplification. For CAD, the annealing temperature during the last cycle was usually set to 51 °C.
Aliquots of all PCR products and negative controls were visualized by electrophoresis through a 1 % agarose gel. In those cases where the PCR yielded multiple bands after adjusting amplification conditions, gel extraction with Omega BioTek D2500-02 kit was used to extract products of the expected size. The extraction was done according to the protocol to a final volume of 30 µL. If the extracted product was too weak (<5 ng/µL) a new PCR using the same primers was employed. In these cases, the described three-step touchdown program was altered to a two-step touchdown program by excluding the first cycle. The annealing temperature in the last cycle was set to 51°C.
Nested PCR was used for one sample that would not amplify for CAD1 using the standard primers (54F+405R). A second PCR with the 122F+909R primer set was then carried out using 2 µL from the first PCR product as a template.
PCR products were purified using ExoSAP-IT® (USB, Cleveland, OH, USA). The purification followed the manufacturer's instruction except for a ten times dilution of the initial ExoSAP-IT solution. Bi-directional sequencing was done by Eurofins MWG Operon (Ebersberg, Germany) with the same primer pair used for amplification.

Table 1, Primer sequences used for amplification of the selected gene segment. F and R in the

Gene segment	Oligo name	Sequence (5'-3')	Reference
CAD1	54F	GTNGTNTTYCARACNGGNATGGT	Moulton and Wiegmann 2004
	405R	GCNGTRTGYTCNGGRTGRAAYTG	Moulton and Wiegmann 2004
	122F	CCACTYATYGGNAAYTATGGNGT	Ekrem unpublished
	909R	AAYYTMAATGAYAAYTCNAAYGARGGA	Ekrem unpublished
CAD4	787F	GGDGTNACNACNGCNTGYTTYGARCC	Moulton and Wiegmann 2004
	1098R	TTNGGNAGYTGNCCNCCCAT	Moulton and Wiegmann 2004
AATS1	A1-92F	TAYCAYCAYACNTTYTTYGARATG	Regier 2008
	A1-244R	ATNCCRCARTCNATRTGYTT	Feng-Yi Su et al. 2008
TPI	TPI-111Fb	GGNAAYTGGAARATGAAYGG	Bertone et al 2008
	TPI-281R	TRNCCNGTNCCDATNGCCCA	Brian Cassel unpublished
PGD	PGD-2F	GATATHGARTAYGGNGAYATGCA	Regier 2008
	PGD-3R	TRTGIGCNCCRAARTARTC	Brian Cassel unpublished

2.1.4. Sequence Editing and Alignment

Contigs from the matching forward and reverse reaction were assembled using DNA Baser v3.2.5 (Heracle Software). This was typically done automatically, followed by manual inspection and trimming of low quality ends. At positions with multiple fluorescent peaks, IUPAC Ambiguity codes were applied. In the cases were only one direction was successfully amplified (occasionally for PGD and TPI), the reliability of the single amplified sequence was evaluated by comparison to chromatograms of assembled forward and backward contigs with high quality values (QV). This was done in order to detect possible ambiguities and sequencing errors.
For each gene segment the corresponding coding sequence (CDS) in *Drosophila melanogaster* was downloaded from Flybase (McQuilton et al. 2012). Intron positions and mutation sites are given with reference to this sequence in nucleotide positions.

Alignment of the sequences was carried out in MEGA 5 (Tamura et al. 2011) using the implemented Clustal W (Thompson et al. 1994). The sequences were first aligned by nucleotides with Clustal W under default settings. The corresponding *D. melanogaster* gene was manually aligned with the other sequences and used to determine reading frame. The *D. melanogaster* sequence was also helpful with detection of introns, which were excised from the alignment using the GT-AG rule (Rogers et al. 1980), since their low quality and highly variable length and content made them difficult to align reliably. Nucleotides were then translated and aligned by amino acids with Clustal W using default settings followed by manual inspection. After the introns were removed and codons were aligned there were only single and double codon induced gaps left in the alignment and no need for further adjustments. Single gene alignments and the concatenated alignment was divided into codon positions and exported as codon

position specific alignments using Mesquite (Maddison 2011). These were later used for model and partitioning testing.

2.2. Choosing genetic loci

In addition to the nuclear primers for CAD1, which was successfully used on *Paratanytarsus* and related genera in Ekrem et al. (2010) and CAD4, primer pairs for 5 new markers, AATS1, AATS2, TPI, PGD and RNA POLI2 were available for phylogenetic analysis on *Paratanytarsus* and closely related outgroups. Of these five, the two most applicable in terms of PCR success, sequencing success and phylogenetic signal were to be included as markers in the final phylogeny along with CAD1 and CAD4. This was considered an acceptable compromise between the time and money available for this thesis, and the need to collect information from independent markers.

In order to test the phylogenetic signal, a subset of 14 specimens were chosen from the total sample set. This was done based on a preliminary phylogeny made using CAD1 and CAD4 sequences, which were the first markers to be sequenced in this project. To include various taxonomic levels, closely related individuals were selected from the *P. austriacus* group in addition to various more distantly related *Paratanytarsus* species and at last 4 species from the different outgroups. Some of these were connected with branches on the CAD tree with relatively low support, so it was of interest to see if any of the new markers could add additional information to those nodes.
To visualize the phylogenetic content of the new gene fragments, likelihood mapping (Strimmer et al. 1997) was performed using Tree-Puzzle 5.2 (Schmidt et al. 2002). This method randomly chooses groups of four samples (quartets) and assigns a likelihood value to each of the three possible unrooted tree topologies, which are then plotted into an equilateral triangle. The resolved topologies with little conflicting data are plotted in the three corners, and indicate the presence of tree-like phylogenetic signal whereas unresolved quartets ends up in the central region. The model applied to each marker was selected using the model selection feature in MEGA 5 under maximum likelihood. The model that generally performed best for all the gene fragments under Bayesian Information Criterion (BIC) (Gideon 1978)) and was available in Tree-Puzzle was the General Time Reversible model with discrete Gamma distribution (GTR+G). This model was therefore applied to all gene fragments, with all model parameters imported from MEGA and 5 discrete gamma categories. Since only 14 sequences were included, the number of possible quartets was low enough to run an exact analysis.
A drawback with likelihood mapping is that the analysis does not give information about the level at which the phylogenetic signal is present. Therefore, in order see how much each marker contributed to resolving the branches that were least well resolved by

CAD1 and CAD4, maximum likelihood (ML), trees with bootstraps values from 1000 replicate runs were made using MEGA 5 (Tamura et al. 2011) and the GTR+G model. Finally, to look for substitution saturation, the observed number of transitions and transversions for each gene fragment were plotted against the genetic GTR corrected distance using DAMBE v5.2.57 (Xia et al. 2001). All parameters were estimated by this program.

2.3. Phylogenetic analysis

2.3.1. Maximum Likelihood analysis

Maximum likelihood (ML) analysis on the final dataset was done using RAxML v. 7.2.6 (Stamatakis 2006) at the Bioportal computer service (http://www.bioportal.uio.no) hosted by the University of Oslo. For the full dataset, 100 independent runs from maximum parsimony starting trees were performed to find the highest scoring tree. Node support was calculated using 1000 nonparametric bootstrap replicates. GTRGAMMA was chosen as substitution model, as it is recommended by the RAxML manual over the more effective GTR-CAT approximation for smaller datasets, and since it produces comparable likelihood values that can be used for model testing. The advice from the manual of not to model invariant sites was also followed. For nucleotides, RAxML only implements GTR and GTR approximations, therefore this substitution model was used for all partitions.

For simplicity during likelihood scoring of the different data partitioning alternatives, a RAxML GUI (Silvestro et al. 2011) with built in options for data partitioning was used. In those cases, the likelihood value was estimated from the best of 10 or fewer independently made trees, as the difference in likelihood between alternative trees was found to be far too low to influence the overall outcome of this test.

Single gene phylogenies and a phylogeny on amino acid sequences were estimated under maximum likelihood in MEGA 5. The best model was selected by the built-in model tester in MEGA, using BIC, as this criteria has been found to select the correct model more often than corrected Akaike Information Criterion (AICc) (Hurvich et al. 1989)), under modeltesting with simulated datasets in MEGA (Tamura et al. 2011). The analysis were run with 5 discrete gamma categories and 1000 bootstrap replicates for confidence estimates.

2.3.2. Bayesian analysis

The Bayesian Metropolis coupled Markov chain Monte Carlo (MCMC) analysis was performed using a parallel version of MrBayes v3.1.2 (Huelsenbeck et al. 2001; Ronquist et al. 2003; Altekar et al. 2004) at the Bioportal computer service. The best scoring

model among those available in MrBayes was chosen from the modeltest in MEGA. Settings for priors and MCMC were left as default. Each of the four MCMC chains were set to run for $5x10^6$ generations with sampling every 100 generations. The first 25 % generations were discarded as burn-in. Mixed partition runs were conducted using the "unlink" and "ratepr = variable" commands to allow parameters to be independently estimated in each partition as recommended by Marshall et al. (2006). Tracer V1.5.0 (Rambaut et al. 2008) was used to estimate the effective sample size (ESS) of each parameter and to plot ln-likelihood values across generations. This information was used together with the standard deviation of split frequencies to ensure chain convergence.

2.3.3. Data Partitioning

It has been demonstrated empirically and by simulations that partitioning data into separate classes, so that model parameters can be unlinked between partitions, can give significant benefits in likelihood scores and nodal support in maximum likelihood analysis (McGuire et al. 2007; Ward et al. 2010) and Bayesian analysis (Nylander et al. 2004; Rota 2011).

A test was therefore set up to explore the effect of different partitioning schemes on the concatenated dataset. The gene-segments were treated in three ways ranging from no partitions where all markers were treated the same, to CAD1 and CAD4 pooled with AATS1 and PGD in separate partitions and all gene-segments treated as a separate partition. Since CAD1 and CAD4 are fragments of the same gene, it was hypothesized that they would benefit from being in the same partition with parameters estimated jointly. The codons were also handled in three ways: no partitioning, third codon as a separate partition and all three codon positions as separate partitions. In total this gave nine partitioning regimes. Due to the time extensive computations involved with sampling likelihood values from the stationary phase of an MCMC run, a reduced partitioning regime was set up for the Bayesian partition model testing based on the results from the maximum likelihood analysis.

To avoid over-parameterization, model selection criteria that impose parameterization penalties were applied to the different partitioning regimes. For the ML based inference, two criteria were used: Akaike Information Criterion (AIC), was calculated as $AIC_i = -2\ln L_i + 2k_i$ where L_i is the likelihood of the data given the model and tree, estimated by RAxML, and k_i is the number of estimated parameters which includes the GTR model, the rate heterogeneity parameter gamma, and nucleotide frequencies. Parameters for branch lengths were not included. The BIC puts a stronger penalization on the number of parameters and was calculated as $BIC_i = -2\ln L_i + k_i * \ln n$, where n in this case was the number of positions in the total alignment.

For the Bayesian analysis, Bayes factor (BF) was used to compare partitioning schemes. This was done through the average harmonic mean of the likelihood values from the stationary face of each MCMC run, using the "sump" command in MrBayes and discarding the burnin. BF was then calculated as the ratio of the likelihood between the two models that was compared, BF = $(-\ln L_i)-(-\ln L_j)$ (Nylander et al. 2004). Some guidelines for the interpretation of BF was presented by Kass et al. (1995). Values in the interval 3-20 gives a positive support of the better model, 20-150 gives a strong support and above this the support is very strong.
A problem with this way of estimating BF is that the harmonic mean estimation is unstable between separate runs (Ronquist et al. 2010) and maybe in favor of parameter-rich models (Lartillot et al. 2006). However it has been found to be a statistically sound method for choosing partitioning strategies in Bayesian phylogenetic inference (Brown et al. 2007).

2.3.4. *Paratanytarsus austriacus / hyperboreus*

To investigate the relationship between the different clusters in the *P. hyperboreus/austriacus* group in more detail, a separate phylogenetic analysis was done on this part of the tree with addition of available mitochondrial COI sequences. This was done to see if more characters could improve the resolution of this part of the tree, and investigate if the COI sequences, which, being mitochondrial differ from nuclear evolution in several ways (e.g. Ballard et al. 2004) would show the same phylogenetic relationship.
A Bayesian analysis was run on the nuclear markers from this group and COI included as a separate partition. Any further partitioning was avoided because of the few characters available for parameter estimation.
MEGA 5 was used to compute genetic Kimura 2-parameter (K2P) distances for the CO1 sequences within this group. Although the validity of the K2P model to estimate distances between closely related COI sequences has been questioned (Srivathsan et al. 2012), this was done to allow distance comparison with other studies.

2.3.5. Biogeography and age

In the dated molecular phylogeny of Chironomidae presented by Cranston et al. (2012), two *Paratanytarsus sp.* (sp. 1 and sp. 2) were included and their common ancestry dated on a BEAST analysis calibrated with the age of known Chironomidae fossils. To see where the time calibrated node separating these two *Paratanytarsus* species would end up in the phylogeny presented here, the CAD1 and CAD4 sequences of these two specimens were downloaded from GenBank and added to the CAD dataset in this study. A separate phylogenetic analysis under ML in MEGA was done in order to estimate the

position of the dated branch point in relation to the other *Paratanytarsus* lineages included here.

3. Results

3.1. Phylogenetic utility of PGD, AATS1 and TPI

3.1.1. Sequencing and alignment

Of the nine primer combinations that were tested, the three new markers, PGD, AATS1 and TPI in addition to CAD1 and CAD4 amplified successfully for all the 14 samples chosen for the marker test. Introns were identified and excised from three alignments; in CAD1 at position 725 an intron of up to 90 base pairs (bp) was removed (reference sequence, GenBank: AAF48639) and for AATS an intron of approximately 60 bp was removed at position 533 (GenBank: AAF52657). For TPI, the sequence quality varied greatly along the amplified region and in most cases only the reverse direction, was successfully sequenced. At position 114 (GenBank: AAN14219), about 80 bp of an especially variable region, possible intron, was removed due to alignment problems. An intron of about 60 bp was also removed at position 187.

Sequence statistics for the markers in the test setup is shown in Table 2, CAD1 and CAD4 are included for comparison. CAD1 and CAD4 had the highest number of variable and parsimony informative sites both in absolute terms and relative to the size of the amplified region. Of the new markers, PGD had the highest number of parsimony informative sites with 189, while AATS and TPI had approximately an equal number of parsimony informative sites with 101 and 98 respectively.

Table 2, Sequence statistics for the gene fragments in the test setup, including nucleotide frequencies, amplified length (bp) and number of variable sites and parsimony informative sites.

Gene	T	C	A	G	Length	Variable sites	Parsimoy informative sites
CAD1	0.31	0.16	0.32	0.22	813	330 (0.41)	218 (0.27)
CAD4	0.30	0.15	0.33	0.22	834	315 (0.38)	228 (0.27)
PGD	0.29	0.20	0.28	0.23	755	258 (0.34)	189 (0.25)
AATS	0.32	0.17	0.27	0.25	431	147 (0.34)	101 (0.23)
TPI	0.29	0.21	0.24	0.26	440	149 (0.34)	98 (0.22)

3.1.2. Phylogenetic signal

Substitution saturation was observed for AATS1, as demonstrated by the relationship between the number of transitions and transversions and genetic distance (Figure 2). Little or no saturation is indicated for the other markers.

As seen in the likelihood mapping (Figure 3) all markers showed a substantial phylogenetic signal. PGD performed slightly better than AATS and TPI with a higher percentage of resolved quartets. AATS and TPI performed about equally.

The ML tree of the concatenated dataset with bootstrap support values from single gene phylogenies is shown in Figure 4 . None of the new gene fragments gave higher node support than CAD1 and CAD4, and especially low support was given from PGD,

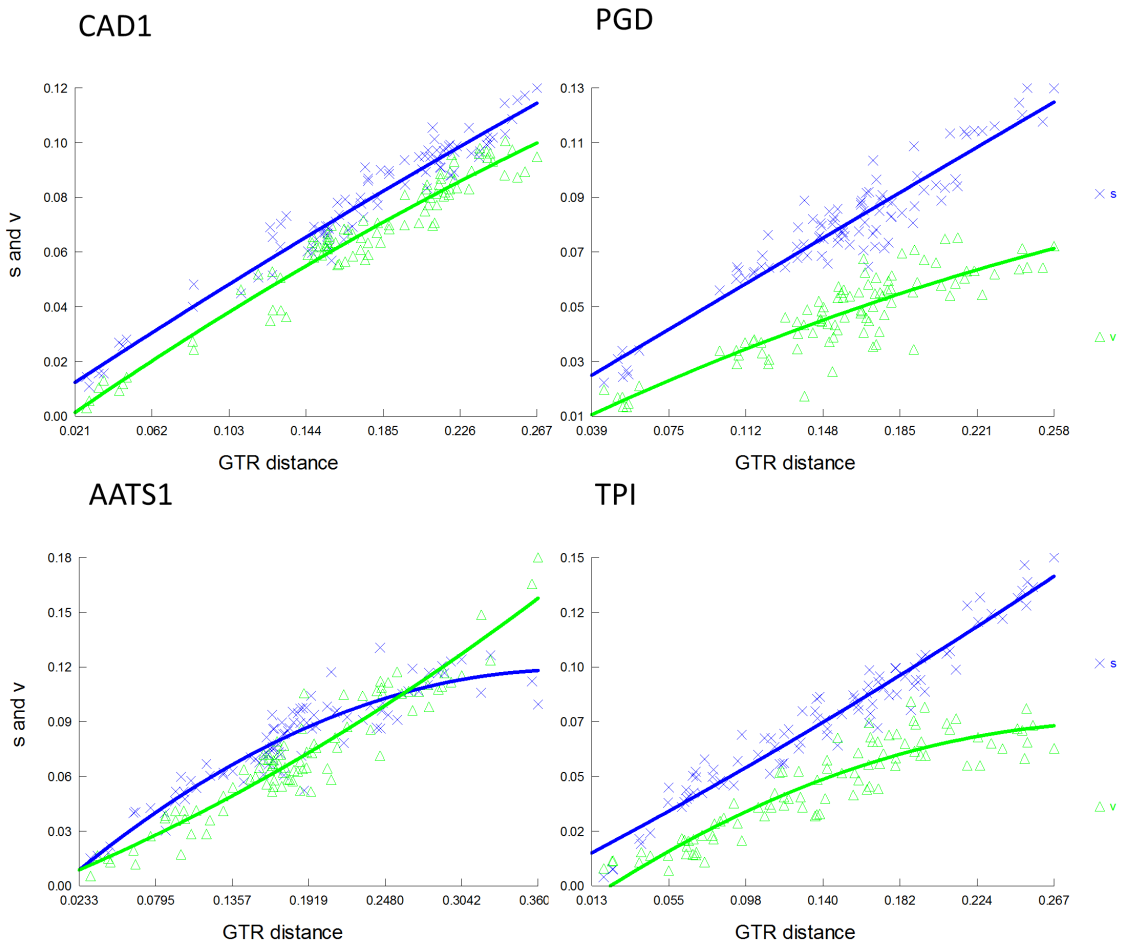


Figure 2 Observed transitions (*s*) and transversions (*v*) against genetic GTR corrected distances for the gene fragments of CAD1, PGD, AATS1 and TPI for the 14 samples in the in the test setup.

AATS1 and TPI to the basic nodes building the backbone of the *Paratanytarsus* tree. Conflicting topologies between individual gene fragments were present due to different arrangements of samples from *P. austriacus* and *P. hyperboreus* in PGD, AATS1 and TPI,

the placement of *P. setosimanus* Goetghebuer, 1933 with respect to the *P. penicillatus* Goetghebuer, 1928 branch in TPI, and the placement of two *Micropsectra* species within *Paratanytarsus* in AATS1.

Since only two new markers were to be included to the total dataset, PGD, which had the highest number of parsimony informative sites and performed best during likelihood mapping, was chosen along with AATS1. TPI performed about equally to AATS1 during likelihood mapping, but was dismissed primarily because of the trouble involved in producing reliable sequences for the alignment, as it contained two possible introns and most often only sequenced in one direction.

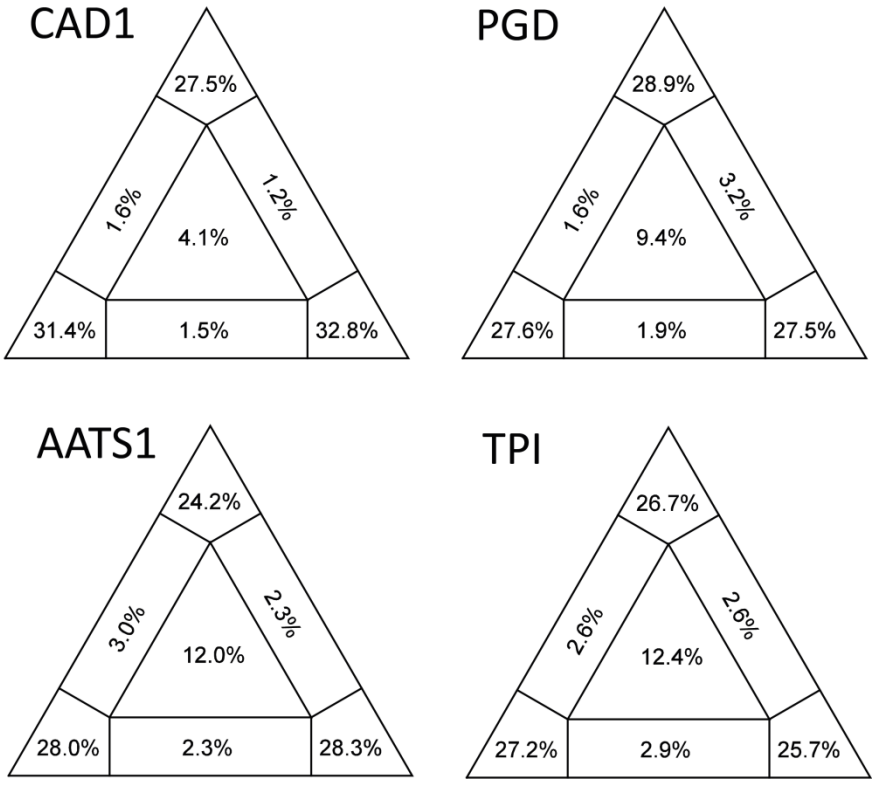


Figure 3, Likelihood mapping analyses represented as triangles of the three new amplified regions, PGD, AATS and TPI with CAD1 as a reference. Corner values show the percentages of resolved phylogenies and the central areas shows the percentage of unresolved branches.

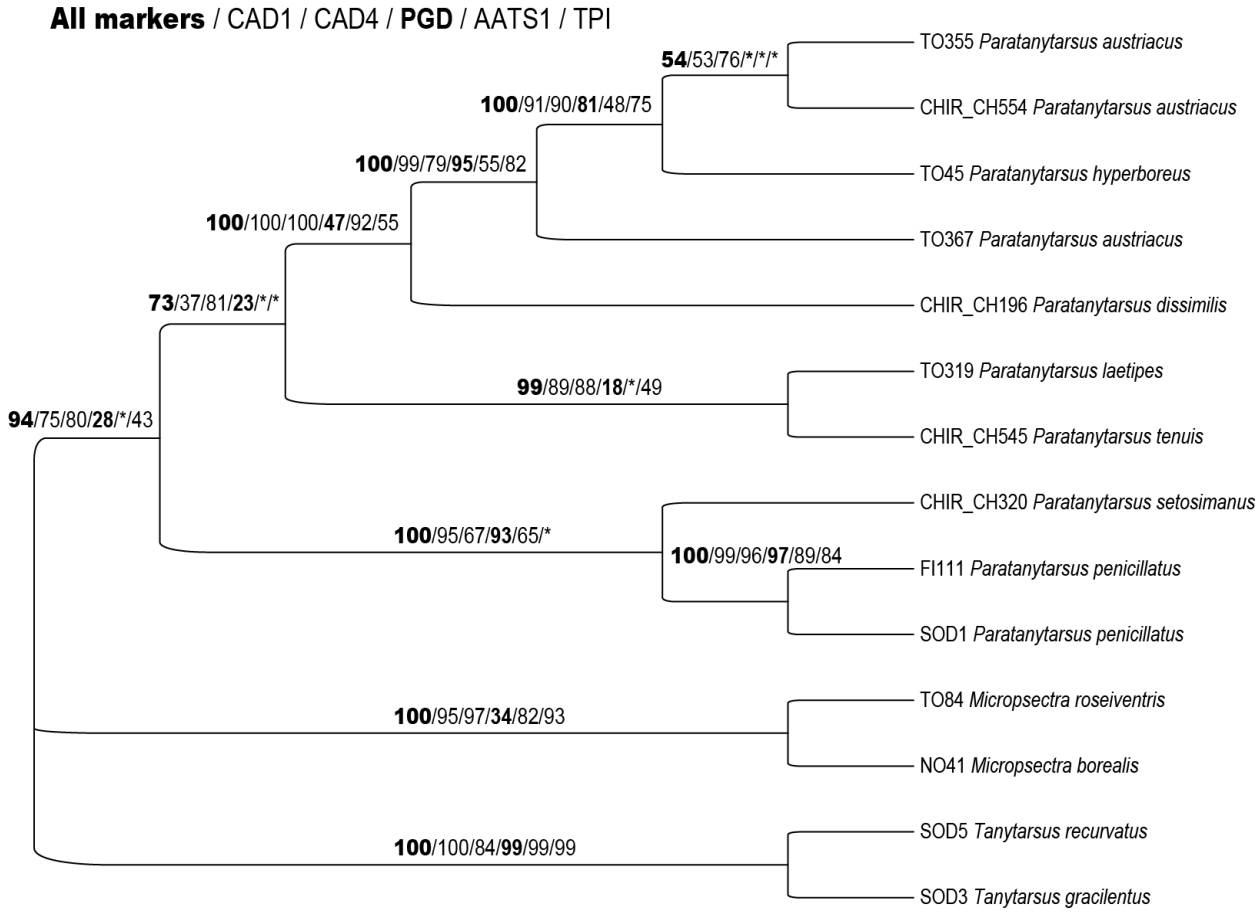


Figure 4, ML tree of the 14 samples in the test setup computed in MEGA 5 under the GTR+G model. The tree is made from a concatenated dataset consisting of CAD1, CAD4, PGD, AATS1 and TPI. Bootstrap values are given for the total dataset and for single gene phylogenies. Disagreements between the concatenated dataset and single gene phylogenies are indicated by a star (*).

3.2. Analysis on the full dataset

3.2.1. DNA sequencing and alignment

Of the total 51 samples selected for this study, all were amplified for AATS1, 49 for CAD1 and CAD4 and 45 for PGD. Details are given in Appendix: Table 2.

After excluding introns, the total concatenated dataset of CAD1, CAD4, AATS1 and PGD consisted of 2862 characters, of which 956 were parsimony informative. Plots of transitions and transversions versus genetic GTR distance (not shown) revealed saturation in codon position three, but no saturation at position one or two.

3.2.2. Data Partitioning and Model Testing

Partitioning of the dataset improved the likelihood score mostly when involving codon positions. Partitioning by gene only differed slightly from the results found with no partitioning. With the ML analysis, AIC favored the most complicated model, which involved partitioning by both gene and codon position, whereas BIC clearly preferred partitioning by codon positions and not by gene. For reasons outlined in the discussion, it was decided to continue with the alternative preferred by BIC, partitioning by all codon positions, for the rest of the ML analysis on the concatenated dataset.

Considering that pooling codon position one and two separately and CAD1 and CAD4 separately differed only slightly from the full partitioning scheme in the ML analysis, these alternatives were ignored for the Bayesian comparison. The favored substitution model used for the Bayesian analysis usually involved the GTR+G model with or without invariant sites, but simpler models were chosen for first position in AATS1 (the SYM model with equal nucleotide frequencies) and third position in CAD1 (HKY+G+I).
After running the Bayesian analysis with the favored model for each partition, it was evident that as for ML, partitioning by codon position resulted in much greater likelihood improvements than partitioning by gene. The highest likelihood was achieved with the most partitioned model, which also got strong support from BF, Table 4. Although the standard deviation of split frequencies was satisfactory for all runs, the ESS value was far too low for the gene + codon partition, indicating convergence problems. This was also evident from parameter plots viewed in Tracer. Subsequently, a second run with $10x10^6$ generations was tried, also providing unsatisfactory convergence. Thus, the codon partitioning scheme was used for the final Bayesian tree.

Table 3, Data partition schemes and corresponding ln likelihood (lnL) scores for the resulting tree topology in RAxML, weighted with AIC and BIC. Numbers in bold represents the best partitioning schemes chosen by the different criteria.

Number of partitions	Parameters	Partitioned by:	lnL	AIC	BIC
1	10	Unpartitioned	-23544	47108	47207
2	20	1.+2. and 3. codon position	-22666	45372	45570
3	30	1., 2., and 3. codon position	-22591	45241	**45538**
3	30	CAD1+4, PGD, AATS1	-23524	47108	47405
4	40	CAD1, CAD4, PGD, AATS1	-23516	47112	47507
6	60	1.+2. and 3. codon position in CAD1+CAD4, PGD, AATS1	-22618	45355	45949
8	80	1.+2. and 3. codon position in CAD1, CAD4, PGD, AATS1	-22604	45368	46160
9	90	1., 2., and 3. codon position in CAD1+CAD4, PGD, AATS1	-22529	45239	46129
12	120	1., 2., and 3. codon position in CAD1, CAD4, PGD, AATS1	**-22497**	**45233**	46420

Table 4, Comparison of likelihood scores, Bayes factor (BF) and convergence statistics among different partitioning schemes. Natural logarithm likelihood is estimated from the average of the harmonic mean of lnL from two parallel runs in MrBayes, Bayes factor is calculated with reference to the partitioning scheme with the highest likelihood score ($-\ln L_i$)-($-\ln L_{best}$) and split deviation is the average standard deviation of split frequencies of two independent runs after $5x10^6$ generations. The effective sample size (ESS) of lnL was calculated in Tracer.

Partitioned by	lnL	Bayes factor	Split deviation	ESS of lnL
Unpartitioned	-23584.39	961.44	0.00406	2984
Gene	-23569.63	946.68	0.00342	2268
Codon	-22718.83	95.88	0.00332	2238
Gene and codon	**-22622.95**	**0**	0.00614	23

3.2.3. Phylogeny of *Paratanytarsus*

The phylogenetic analysis under ML and Bayesian inference produced close to identical trees and there were no conflicting topologies. The dataset appeared stable; only small topological differences were present among the different partitioning schemes in the ML and Bayesian analysis and the amino acid tree.

As a genus, *Paratanytarsus* came out as well supported monophyletic group (bootstrap value = 88, posterior probability = 1). Five species came out paraphyletic, *Paratanytarsus natvigi*, *P. austriacus* and *P. hyperboreus. P. dissimilis* Johannsen, 1905*,* and *P. tenuis* Meigen, 1830*,* the last two with respect to samples from the Nearctic region. One *P. penicillatus* (FI111), separated from the other *P. penicillatus*, including those sampled at the same location.

The *Paratanytarsus sp.* from Australia came out well-embedded in the tree, separated from the other species by a long branch. The amount of genetic difference across the Atlantic varies between the species. For example, *P. setosimanus* and a northern clade of *P. austriacus/*hyperboreus shows little difference between samples from Canada and Norway, whereas in *P. austriacus, P. dissimilis* and *P. tenuis*, there are obvious genetic differences.

3.2.4. Single gene trees

Comparison of single gene trees confirmed the monophyly of *Paratanytarsus* based on CAD1, CAD4 and PGD, with bootstrap supports of 64, 44 and 38. *Paratanytarsus* was not monophyletic based on AATS, but the bootstrap support against monophyly in this marker was low. The placement of *P. abiskoensis* Reiss et al., 1981*, P. bituberculatus* Edwards, 1929*, P. laetipes* Zetterstedt, 1850 and *P. intricatus* Goetghebuer 1921, is least supported in all the trees. Nevertheless, they end up at about the same place in all markers except CAD1, where they are found more scattered, although with low support. In conclusion, the trees only differ in regions with low support, and there are no major disagreements.

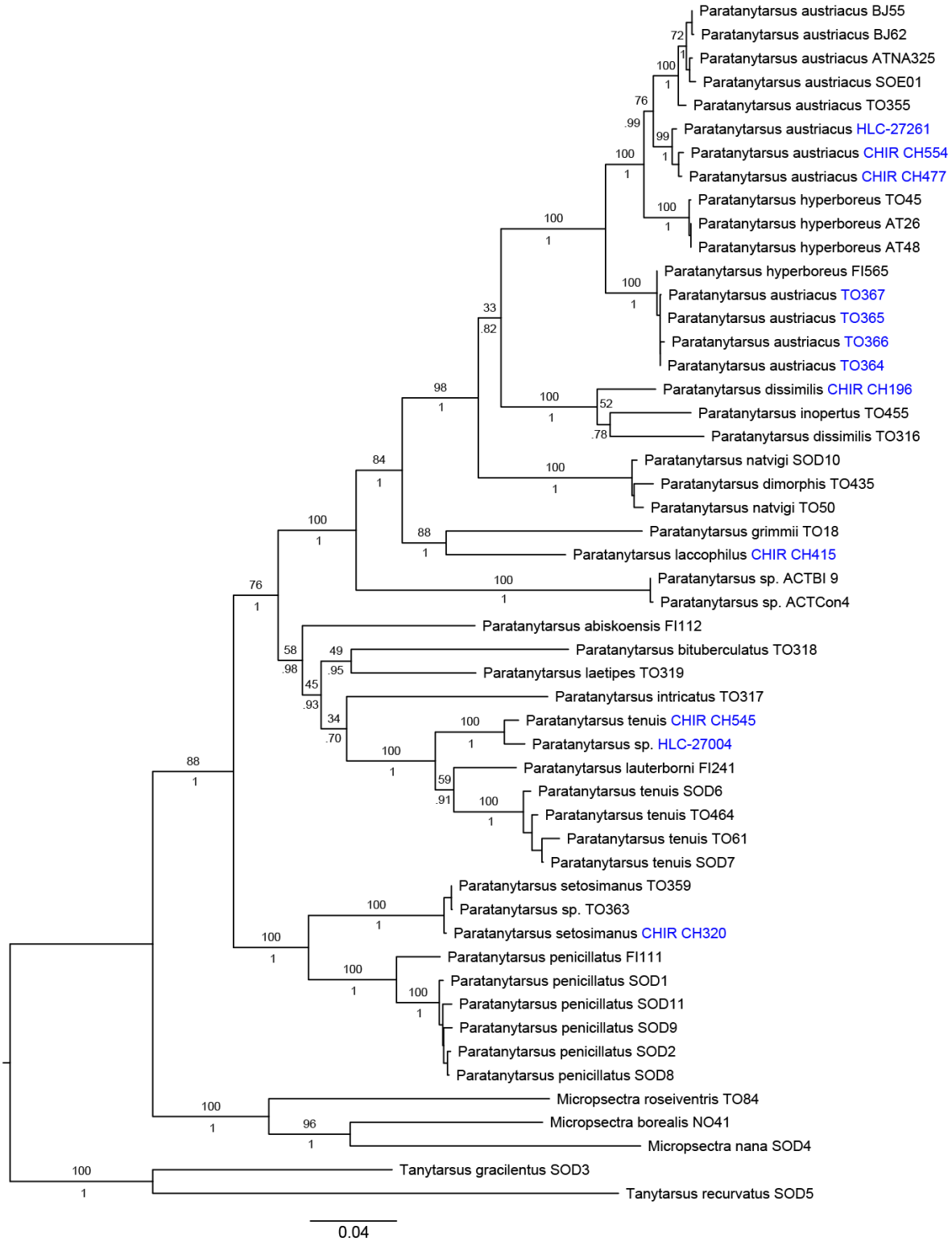


Figure 5, Maximum likelihood tree from the analysis on the concatenated dataset. Bootstrap values are given above the branches and Bayesian posterior probabilities are given below. Sample ID in blue are from the Nearctic region.

3.2.5. P. austriacus/hyperboreus group

A separate phylogeny on the CO1 mitochondrial marker for the samples of *P. austriacus* and *P. hyperboreus* revealed the same phylogenetic pattern as for all the nuclear markers (Figure 6), K2P CO1 distances between different clusters in this tree were in the range of 15.0 %, (between the two clusters represented by TO364 and AT48), to 9.3 % (between the clusters represented by ATNA325 and CHIR_CH477). The maintenance of stable genetic clusters in all the investigated markers, combined with a probable geographic overlap between three of the clusters involved, indicates that more than two species is present within this group. Morphologically, no consistent characters have been found to separate the different groups, but they have yet to be studied in detail.

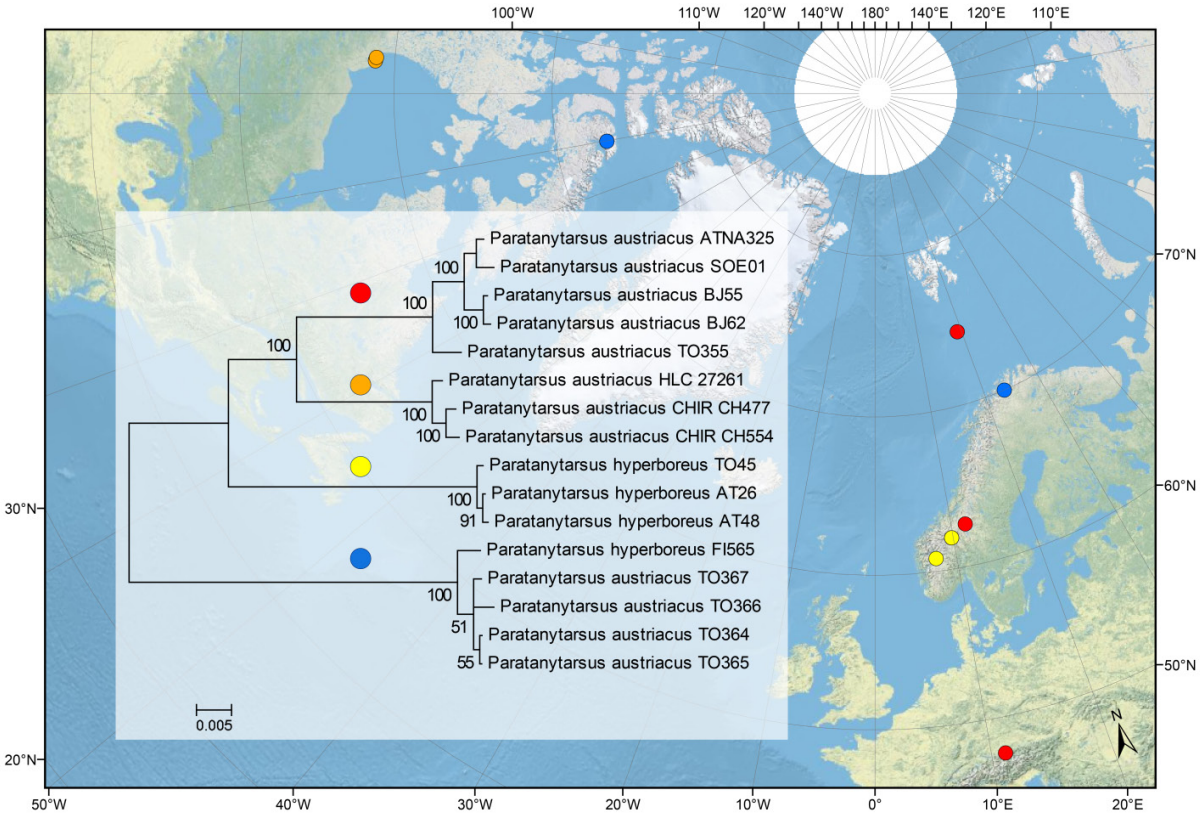


Figure 6, Bayesian tree on the combined dataset of the mitochondrial marker COI and the nuclear markers CAD1, CAD4, PGD and AATS for the taxa in the *P. austriacus* and *P. hyperboreus* group. The colors on the phylogeny indicates the sampled location. The data matrix was partitioned in two, one nuclear DNA partition analyzed under the GTR+I model, and a mtDNA partition analyzed under GTR+G.

3.2.6. Biogeography and age

The *Paratanytarsus* sequences from Cranston et al. (2012) ended up close to the node of the Australian species in this study (ACTBI 9, ACTCon4). The first of the added species, "*Paratanytarsus sp. 1*" shared node with ACTBI 9/ACTCon4, but was well separated by a long branch. "*Paratanytarsus sp. 2*" grouped close to *P. laccophilus* Edwards, 1929 (CHIR_CH415) right next. Therefore, the age of 50 million years (95 % confidence interval: ± 20 million years) estimated for the *Paratanytarsus* branch in Cranston et al. (2012) could be representative for the branching point between the Australian species and *P. laccophilus* in the presented phylogeny.

4. Discussion

4.1. Phylogenetic signal and data partitioning

For the deepest nodes in the phylogeny, CAD1 and CAD4 provided most support, while the contribution from the other markers was lower and more variable. Therefore, it appears that most of the backbone in the *Paratanytarsus* tree has been based upon information from CAD. AATS1 had the highest genetic distances and the highest substitution saturation and is apparently most suitable for answering lower-level phylogenetic questions in Chironomidae. TPI showed promising characteristics in terms of substitution saturation, however other TPI primers than those used here should be developed for this part of the Diptera tree. PGD had little substitution saturation and more characters than AATS1, but contained surprisingly little information about some of the distant relationships within *Paratanytarsus*. This picture is in agreement with Bertone et al. (2008) who found CAD4 superior to PGD and TPI in reconstructing deep nodes in Diptera. Considering that CAD1 and CAD4 also had strong resolution at the lower-levels, the findings presented here adds to the literature describing desirable properties of CAD for reconstructing Diptera phylogenies (Moulton et al. 2004; Winterton et al. 2007; Winkler et al. 2009; Ekrem et al. 2010).

For the final phylogenetic analysis, all gene fragments were pooled and partitioned by codon position. Additional partitioning by gene further increased the likelihood score, however this gave much smaller improvements relative to the likelihood score already obtained. This indicates that there was less differences between gene fragments than between codon positions within one gene fragment.

When new partitions are introduced, the number of free parameters relative to the amount of data rises rapidly; correspondingly, so does the variance in parameter estimates, which could potentially result in misleading parameter estimates (Brown et al. 2007). Since the difference in likelihood between the gene + codon partitioning scheme and the codon partitioning scheme was relatively small, it was decided that the additional partitioning did not increase the likelihood enough to justify the increased parameterization. Thus, the alternative preferred by BIC, partitioning by all codon positions, was used.

4.2. Monophyly of Paratanytarsus

The goal of modern taxonomy is to form groups that are defined by shared ancestry. Central to this is the establishment of monophyletic groups, which by definition should contain all descendants of a common ancestor, without involving any outside lineages. There might, however, be earlier ancestors that define a more inclusive monophyletic group.

The monophyly of *Paratanytarsus* was questioned in Ekrem et al. (2010), where the genus only came out as monophyletic with one of the included markers, CAD1 and the concatenated dataset. Importantly, though, CAD1 was found to be superior to the other markers for the included taxa (Ekrem et al. 2010).

Kravtsova et al. (2010) investigated phylogenetic relationships between *P. baikalensis* and various *Paratanytarsus* and *Micropsectra* taxa using the COI mtDNA marker. The result showed a close relationship between *P. baikalensis* Chernovskij, 1949*, P. austriacus* and *P. hyperboreus*, a finding that might well be valid, given the comparatively short genetic distances between these species. On the other hand, the inferred higher phylogeny where *Paratanytarsus* is paraphyletic with regard to *Micropsectra* (Kravtsova et al. 2010) is certainly questionable, considering that COI has been demonstrated to provide a poor phylogenetic signal at genus-level in Chironomidae (Ekrem et al. 2007; Ekrem et al. 2010).

With the exception of AATS1, all markers in this study, including TPI from the test setup, supported the monophyly of *Paratanytarsus.* For AATS1, *Paratanytarsus* is found to be paraphyletic with regard to *Micropsectra*, a result that also is consistent when removing the variable third codon position. Although this result is surprising, it probably should not be given strong emphasis due to the low branch support provided in this scenario, as well as the generally low resolution of this marker at the basal nodes of the tree. Also, considering *Micropsectra* ends up right next to the Australian taxa, ACTBI 9/ACTCon4, which has the longest species-specific branch length in the phylogeny, the result might be a consequence of long-branch attraction.

Although the monophyly of *Paratanytarsus* was given high support from the concatenated dataset in this study, species not adequately sampled or not yet discovered could certainly exist, which may affect the results seen here. One candidate is the morphologically deviating species *Paratanytarsus longistilus* recently described from the Nearctic region, that possess a set of characters which deviates from the former description of *Paratanytarsus* as mentioned in the introduction (Bolton et al. 2010).

4.3. Phylogenetic relationships

Relationships within *Paratanytarsus* have been suggested based on pupal and adult male morphology (Pinder et al. 1986; Cranston et al. 1989). The conclusions drawn from these two life stages have been congruent. The later of the two studies included a greater number of species, resulting in a total of eight species groups. One of these species groups consists of *P. setosimanus* + *P. penicillatus,* a relationship well supported by the molecular phylogeny (100 % bootstrap). Less support is given to the two species groups *P. bituberculatus* + *P. lauterborni* Kieffer, 1909, and *P. tenuis* + *P. intricatus* + *P. laetipes*. The discrepancy is caused by the presence of *P. lauterborni* between samples of *P. tenuis* from Norway and Canada, not adjacent to *P. bituberculatus* as the morphological species groups implied. Apart from *P. lauterborni*, all the species in this part of the tree, branching off from *P. abiskoensis*, are well separated by long branches. However, low branch support indicates difficulties involved in resolving their inner relationship. Moving up in the tree, the Australian species, represented by two spesimens, are deeply embedded in the tree without any close relatives. The parthenogenic and widely distributed *P. grimmii* is grouped with *P. laccophilus* from Canada, although the pair are clearly diverged.
P. natvigi and *P. dimorphis* appear very closely related, with *P. dissimilis* and *P. inopertus* as distant neighbors, a view which is in agreement with earlier proposals as they (*P. dissimilis* not included) have been placed in the same species group (Pinder et al. 1986; Cranston et al. 1989).

4.4. P. austriacus / hyperboreus

Paratanytarsus austriacus was described by Kieffer (1924), from material collected in Lunzer Mittersee, Austria. Later, Brundin (1949) described *P. hyperboreus* based on samples collected in Jämtland, Sweden. The two species are morphologically similar and are separated as adult males by the shape of minor characters of the male hypopygium such as the length and form of the median volsella and the circumference of superior volsella (Reiss et al. 1981).

The phylogenetic analysis of individuals identified to both *P. austriacus* and *P. hyperboreus* revealed four distinct genetic clusters. Among these was a specimen from southern Bavaria, close to the type locality of *P. austriacus* (Figure 6, red cluster). Taking this factor into account, both morphology and geography indicate that this cluster represents the "true" *P. austriacus*. The nearest cluster to this "true" *P.* austriacus is represented by a cluster of specimens collected in Manitoba, Canada (Figure 6, orange cluster). Whether this group represents a Nearctic sister to the "true" European *P. austriacus*, or deserves to be ranked as a separate species should be investigated by broader sampling in this region, which could clarify if other clusters of *P. austriacus* are present.

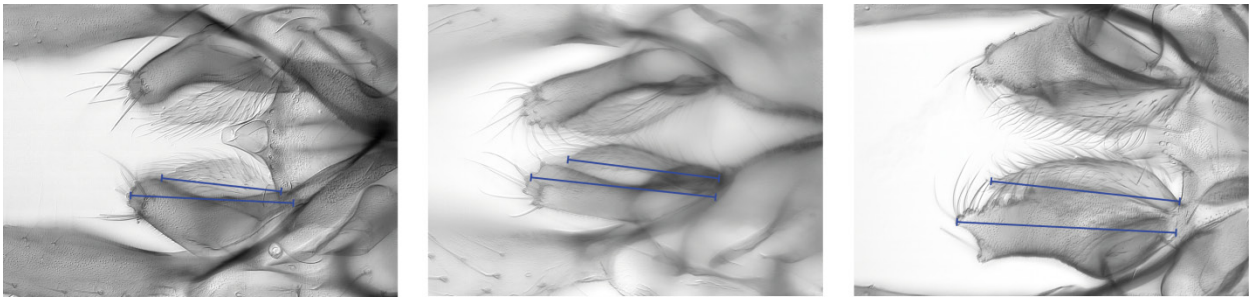
In Norway there are three distinct branches, one represented by a sample identified as *P. hyperboreus* from northern Norway (Figure 6, blue cluster), and two represented by samples collected in central Norway (Figure 6, yellow cluster). The K2P distance of partial COI sequences from the Norwegian samples were in the range of 15.0 % (between *P. hyperboreus* FI565, blue cluster and *P. hyperboreus* TO45, yellow cluster), to 13.3 % (between *P. hyperboreus* AT26, yellow cluster, and *P. austriacus* ATNA325, red cluster). In comparison, the average intraspecific K2P distance for chironomids found in Ekrem et al (2007), was 16.2 % with a variation from 5.1 % to 25.2 %.

For recently derived species, incomplete lineage sorting can potentially complicate species delimitations due to gene trees that differ in topology from locus to locus (Knowles et al. 2007). When such incongruent genes are combined in a concatenated datasets, the result can give misleading inferences about the history of divergence (Kubatko et al. 2007). Moreover, mitochondrial genes may be especially prone to conflicts with nuclear genes, given differences in traits such as ploidy, mode of inheritance, degree of recombination, effective population size and mutation rate (Scheffler 1999). In light of this, the molecular dataset of the *P. austriacus* and *P. hyperboreus* taxa, which displayed the same stable tree topology for all three nuclear genes and the mitochondrial COI gene, strengthens the hypothesis that the gene trees are representative of the evolutionary history of this species group.

In the process of delimiting species, one increasingly used criterion is the phylogenetic species concept (Cracraft 1983), under which a species is considered as the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent. For the molecular dataset, this concept clearly suggests the presence of one new species in Norway, in addition to the two already described. The phylogenetic species concept does, however, evade the problem of reproductive isolation involved with the biological species concept (Mayr 1963). Given the relatively short geographic distances between the three clusters sampled within Norway, and the widespread distribution of one of the taxa, the "true *P. austriacus*", which probably

overlaps with the two other Norwegian clusters, or has done so in biologically recent times, suggests that they also are species according to the biological species concept, and that reproductive barriers exists.

Although time constraints have not permitted detailed morphological analysis of all of the specimens, currently used diagnostic features have been examined. One of these characters is the length of the median volsella which is described as greater than 90 µm in *P. austriacus* and shorter than 70 µm in *P. hyperboreus*, and its relative size compared to the inferior volsella; in *P. austriacus* these two structures nearly reach the tip of each other (Reiss et al. 1981).
Length measurements of the median volsella of the specimens included in this study revealed a large overlap in the ranges between different genetic clades, indicating that this character is not as useful for species recognition as previously assumed (Figure 7). While size comparisons of characters such as the median and inferior volsella may be important diagnostic tools in some cases, this may not always be a reliable taxanomic tool. For example, certain species within the closely related genus of *Micropsectra* may be differentiated based on these types of minute differences. On the other hand, similar lines of evidence used to differentiate other morphologically similar species within the same genus have become blurred as the number of species and localities sampled increases (Anderson, personal comm).
Accurate species delimitation is important, but it is also important to provide morphological characters to diagnose distinct species, in order to allow easier integration in biodiversity studies or in biomonitoring. Although speciation is not always accompanied by morphological change (Bickford et al. 2007), the genetic clusters identified here should be used as a basis for a more thorough morphological examination precluding a possible new species description.



TO365 TO45 BJ55

Figure 7, Relationship between median and inferior volsella in *Paratanytarsus austriacus/hyperboreus* among representatives from three of the separate clusters in Figure 6.

4.5. Biogeography

The available dataset is of limited value for a comprehensive analysis of the historical biogeography of *Paratanytarsus*. Most of the species are sampled from a narrow range in Europe and Arctic Canada, with one additional Australian species represented by two individuals. However, some comments may still be of value.

The dated node from Cranston et al. (2012) ended up close to the basic node of the Australian *Paratanytarsus* in this dataset. Therefore, the estimated 50 million year age (95 % confidence interval: ± 20 million years) of the Australian branch from Cranston et al. (2012) is believed to be a realistic estimation for this node in the phylogeny of *Paratanytarsus* presented here.

In this time period, the early Eocene, the earth was quit similar to its modern day appearance, with respect to land bridges and isolation of continents (Cox et al. 2005). If the idea of a northern hemisphere origin of *Paratanytarsus* presented by Säwedal (1982) is correct, it is likely that *Paratanytarsus* is present in Australian due to dispersal rather than vicariance. Chironomids are weak fliers thus are most likely spread passively by wind and air currents. There are reports of chironomids collected over oceans (Cheng et al. 1978) and at high altitudes (Glick 1960), indicating that their passive migration capacity perhaps should not be underestimated. This is also likely in the light of the low genetic variation found between some samples of the same species separated by the Atlantic ocean in this study, and the colonization of the isolated arctic island Bear Island by *P. austriacus* after the last glaciation.

Various analytical methods used to recognize ancestral origins and biogeographical patterns exist currently, some of which include measures of species richness and phylogenetic diversity (Cranston 1995). However, inconsistencies among different regions is often caused by unequal mapping efforts and sampling intensities. For *Paratanytarsus*, many species are known from the western Palearctic, but much less are known from the Australasian region, and very few from the Neotropical and Oriental regions. These are also the areas where the chironomid fauna has been least thoroughly studied (Ashe et al. 1987), which makes accurate assessments about species richness difficult.

Concerning morphology, none of the Australian *Paratanytarsus* species included here or in Cranston et al. (2012) were identified to species because all specimens were larvae and could not be associated to described pupae or adults. The drawings of described Australian *Paratanytarsus* species (Glover 1973) indicate however, that Australian *Paratanytarsus*, do not constitute a separate group that have evolved in isolation after a possible colonization event. Rather, each species is more similar to species from the northern hemisphere than they are to each other. Looking at a broader spectrum of the

chironomid fauna, Brundin (1966) found a clear pattern of bipolar distribution in Chironomidae species groups, a pattern also represented by cold adapted species. This distribution pattern seems likely for *Paratanytarsus,* as the genus appears to be more diverse in northern and southern temperate regions.
The phylogeny presented here does not dispute the theory that *Paratanytarsus* has a Holarctic origin and that southern distributions are due to secondary dispersal (Säwedal 1982). It is, however, believed that the geographic sampling is too limited to give any further details or to test the theory. For this to be sensible, more *Paratanytarsus* species and related outgroup taxa, particularly from the Australasian, Oriental and Neotropical regions should have to be included.
When concerning trans-Atlantic relationships, the presence of deeply divergent lineages between species identified to *P. tenuis* and *P. dissimilis* indicates separate populations of what might be as of yet undescribed Nerarctic species. Similar patterns have recently been found in the related genus *Micropsectra* (Anderson et al. pers comm.).
Two species in particular show very little transatlantic genetic divergence. If this is due to current gene flow, it could suggests a wider circum-Holarctic distribution, which for many groups of organisms is most likely if the species is restricted to the boreal or arctic zones (Danks 1981). Interestingly, the two species in the phylogeny that shows the least intraspesific variation across the Atlantic are the *P. setosimanus* and the northern clade of *P. hyperboreus/austriacus* (e.g. FI565, TO367). Assuming for now that the latter group represents the true *P. hyperboreus*, this species and *P. setosimanus* are the two most northerly and high altitude distributed species of *Paratanytarsus* (Reiss et al. 1981).

5. Conclusion

While the number of taxa included and the genes sampled in a study of this nature could always be increased to enhanced, the results presented here contributes to the field of Chironomidae research and Dipteran systematic by providing the most detailed phylogenetic analysis of *Paratanytarsus* so far. This work has revealed four separate genetic clusters within the two species *P. austriacus/hyperboreus* , and one new species is likely to exist in Norway. In addition, substantial trans-Atlantic genetic divergence suggests two undescribed Nearctic species. So far the monophyly of *Paratanytarsus* is well supported, although wider sampling of species might change this picture in the future. Of special interest in that respect would be to include more material from Australia and the other biogeographical regions which have not been included here. This could throw much needed light on some of the biogeographical questions that have been raised here, including the idea of bipolar migrations.

References

Altekar, G., S. Dwarkadas, J. P. Huelsenbeck and F. Ronquist (2004). "Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference." *Bioinformatics* **20**(3): 407-415.

Ashe, P. (1983). "A catalogue of chironomid genera and subgenera of the world including synonyms (Diptera: Chironomidae)." *Entomologica Scandinavica Supplement* **20**: 1-68.

Ashe, P., D. A. Murray and F. Reiss (1987). "The zoogeographical distribution of Chironomidae (Insecta : Diptera)." *Annales de Limnologie - International Journal of Limnology* **23**(01): 27-60.

Ballard, J. W. O. and M. C. Whitlock (2004). "The incomplete natural history of mitochondria." *Molecular Ecology* **13**(4): 729-744.

Bause, E. (1913). "Preprint of Bause E. 1914: Die Metamorphose der Gattung *Tanytarsus* und einiger verwandter Tendipedidenarten. Ein Beitrag zur Systematik der Tendipediden." *Archiv für Hydrobiologie Supplement* **2**: 1-139.

Bertone, M. A., G. W. Courtney and B. M. Wiegmann (2008). "Phylogenetics and temporal diversification of the earliest true flies (Insecta: Diptera) based on multiple nuclear genes." *Systematic Entomology* **33**(4): 668-687.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram and I. Das (2007). "Cryptic species as a window on diversity and conservation." *Trends in Ecology & Evolution* **22**(3): 148-155.

Bolton, M. J., T. Ekrem, J. E. Sublette and M. F. Sublette (2010). "A new species of *Paratanytarsus* Thienemann and Bause (Diptera: Chironomidae) with unusual larval and adult male morphology
" *Proceedings of the XV International Symposium on Chironomidae*: 262-271.

Brown, J. M. and A. R. Lemmon (2007). "The importance of data partitioning and the utility of bayes factors in Bayesian phylogenetics." *Systematic Biology* **56**(4): 643-655.

Brundin, L. (1947). "Zur Kenntnis der schwedischen Chironomiden." *Arkiv för Zoologi* **39 A**(3): 1-95.

Brundin, L. (1949). Chironomiden und andere Bodentiere der südschwedischen Urgebirgsseen. Ein Beitrag zur Kenntnis der bodenfaunistischen Charakterzüge schwedischer oligotropher Seen. Lund, Institute of freshwater research, Drottningholm

Brundin, L. (1966). Transantarctic relationships and their significance, as evidenced by chironomid midges. With a monograph of the subfamilies Podonominae and Aphroteniinae and the austral Heptagyiae. Stockholm, Almquist & Wiksell

Carew, M. E., V. Pettigrove, R. L. Cox and A. A. Hoffmann (2007). "DNA identification of urban Tanytarsini chironomids (Diptera : Chironomidae)." *Journal of the North American Benthological Society* **26**(4): 587-600.

Cheng, L. and M. C. Birch (1978). "Insect flotsam - Unstudied marine resource." *Ecological Entomology* **3**(2): 87-97.

Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, et al. (2007). "Evolution of genes and genomes on the Drosophila phylogeny." *Nature* **450**(7167): 203-218.

Cox, C. B. and P. D. Moore (2005). Biogeography: an ecological and evolutionary approach. Malden, Mass., Blackwell

Cracraft, J. (1983). "Species concepts and speciation analysis." *Current Ornithology* **1**: 159-187.

Cranston, P. S. (1995). The Chironomidae. Biology and ecology of non-biting midges, "Biogeography". London, Chapman & Hall. 70

Cranston, P. S. (1995). The Chironomidae. Biology and ecology of non-biting midges, "Introduction". London, Chapman & Hall. X - 7

Cranston, P. S., M. E. Dillon, L. C. V. Pinder and F. Reiss (1989). The adult males of Chironominae (Diptera, Chironomidae) of the Holarctic region. Keys and diagnoses. Chironomidae of the Holarctic region. Keys and diagnoses. Part 3. Adult males. T. Wiederholm, Entomological Scandinavica Supplement. **34:** 353-502.

Cranston, P. S., D. H. D. Edward and L. G. Cook (2002). "New status, distribution records and phylogeny for Australian mandibulate Chironomidae (Diptera)." *Australian Journal of Zoology* **41**: 357-366.

Cranston, P. S., N. B. Hardy and G. E. Morse (2012). "A dated molecular phylogeny for the Chironomidae (Diptera)." *Systematic Entomology* **37**(1): 172-188.

Cranston, P. S., N. B. Hardy, G. E. Morse, L. Puslednik and S. R. McCluen (2010). "When molecules and morphology concur: the 'Gondwanan' midges (Diptera: Chironomidae)." *Systematic Entomology* **35**(4): 636-648.

Danforth, B. N., C. P. Lin and J. Fang (2005). "How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns?" *Systematic Entomology* **30**(4): 549-562.

Danks, H. V. (1981). Arctic arthropods: a review of systematics and ecology with particular reference to the North American fauna. Ottawa, Entomological Society Canada

Ekrem, T., E. Stur and P. D. N. Hebert (2010). "Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding." *Organisms Diversity & Evolution* **10**(5): 397-408.

Ekrem, T. and E. Willassen (2004). "Exploring Tanytarsini relationships (Diptera: Chironomidae) using mitochondrial COII gene sequences." *Insect Systematics & Evolution* **35**: 263-276.

Ekrem, T., E. Willassen and E. Stur (2007). "A comprehensive DNA sequence library is essential for identification with DNA barcodes." *Molecular Phylogenetics and Evolution* **43**(2): 530-542.

Ekrem, T., E. Willassen and E. Stur (2010). "Phylogenetic utility of five genes for dipteran phylogeny: A test case in the Chironomidae leads to generic synonymies." *Molecular Phylogenetics and Evolution* **57**(2): 561-571.

Gideon, S. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics* **6**(2): 461-464.

Gilka, W. and N. Jazdzewska (2010). "A systematic review of the genus *Parapsectra* Reiss (Diptera: Chironomidae: Tanytarsini) with description of a new species from Poland." *Zootaxa*(2350): 1-21.

Glick, C. A. (1960). "Collecting insects by airplane, with special reference to the dispersal of the potato leafhopper." *Technical bulletin of the United States Department of Agriculture* **1222**(16).

Glover, B. (1973). "The Tanytarsini (Diptera: Chironomidae) of Australia." *Australian Journal of Zoology Supplement* **23**: 403-478.

Guryev, V., I. Makarevitch, A. Blinov and J. Martin (2001). "Phylogeny of the genus *Chironomus* (Diptera) inferred from DNA sequences of mitochondrial cytochrome b and cytochrome oxidase I." *Molecular Phylogenetics and Evolution* **19**(1): 9-21.

Hebert, P. D. N., A. Cywinska, S. L. Ball and J. R. deWaard (2003). "Biological identifications through DNA barcodes." *Proceedings of the Royal Society London, B* **270**(1512): 313-321.

Huang, D. W. and L. N. Cheng (2011). "The flightless marine midge *Pontomyia* (Diptera: Chironomidae): ecology, distribution, and molecular phylogeny." *Zoological Journal of the Linnean Society* **162**(2): 443-456.

Huelsenbeck, J. P. and F. Ronquist (2001). "MrBayes: Bayesian inference of phylogeny." *Bioinformatics* **17**: 754-755.

Hurvich, C. M. and C.-L. Tsai (1989). "Regression and Time Series Model Selection in Small Samples." *Biometrika* **76**(2): 297-307.

Kass, R. E. and A. E. Raftery (1995). "Bayes Factor." *Journal of the American Statistical Association* **90**(430): 773-795.

Kieffer, J. J. (1924). "Chironomides nouveaux ou rares de l'Europe Centrale." *Bulletin de la Société d'Histoire naturelle de Metz* **30**: 11-110.

Knowles, L. L. and B. C. Carstens (2007). "Delimiting species without monophyletic gene trees." *Systematic Biology* **56**(6): 887-895.

Kravtsova, L., O. Mayer and D. Shcherbakov (2010). "Phylogenetic relationships of the Baikal endemic *Paratanytarsus baicalensis* (Tshern.) with representatives of genera *Paratanytarsus* Thien. et Bause and *Micropsectra* Kieff. (Diptera, Chironomidae)." *Moscow University Biological Sciences Bulletin* **65**(4): 177-180.

Krosch, M. N., A. M. Baker, P. B. Mather and P. S. Cranston (2011). "Systematics and biogeography of the Gondwanan Orthocladiinae (Diptera: Chironomidae)." *Molecular Phylogenetics and Evolution* **59**(2): 458-468.

Kubatko, L. S. and J. H. Degnan (2007). "Inconsistency of phylogenetic estimates from concatenated data under coalescence." *Systematic Biology* **56**(1): 17-24.

Langton, P. H., P. S. Cranston and P. D. Armitage (1988). "The parthenogenetic midge of water supply systems, *Paratanytarsus grimmii* (Schneider) (Diptera: Chironomidae)." *Bulletin of Entomological Research* **78**: 317-328.

Lartillot, N. and H. Philippe (2006). "Computing Bayes factors using thermodynamic integration." *Systematic Biology* **55**(2): 195-207.

Maddison, W. P. a. D. R. M. (2011). "Mesquite: a modular system for evolutionary analysis. Version 2.74 http://mesquiteproject.org."

Makarevich, I., E. Berezikov, V. Guryev and A. Blinov (2000). "Molecular phylogeny of the *Chironomus* genus deduced from nucleotide sequences of two nuclear genes *ssp160* and the globin 2b gene." *Molecular Biology* **34**(4): 606-612.

Marshall, D. C., C. Simon and T. R. Buckley (2006). "Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors." *Systematic Biology* **55**(6): 993-1003.

Martin, J., A. Blinov, E. Alieva and K. Hirabayashi (2007). A molecular phylogenetic investigation of the genera closely related to *Chironomus* Meigen (Diptera: Chironomidae). Contributions to the Systematics and Ecology of Aquatic Diptera: A Tribute to Ole A. Sæther. T. Andersen. Columbus, Ohio, The Caddis Press**:** 193-203.

Martin, J., V. Guryev, A. Blinov and D. H. D. Edward (2002). "A molecular assessment of the extent of variation and dispersal between Australian populations of the genus *Archaeochlus* Brundin (Diptera : Chironomidae)." *Invertebrate Systematics* **16**(4): 599-603.

Mayr, E. (1963). Animal species and evolution. Cambridge, Mass., The Belknap Press of Havard University Press

McGuire, J. A., C. C. Witt, D. L. Altshuler and J. V. Remsen (2007). "Phylogenetic systematics and biogeography of hummingbirds: Bayesian and maximum likelihood analyses of partitioned data and selection of an appropriate partitioning strategy." *Systematic Biology* **56**(5): 837-856.

McQuilton, P., S. E. St Pierre, J. Thurmond and C. FlyBase (2012). "FlyBase 101-the basics of navigating FlyBase." *Nucleic Acids Research* **40**(D1): D706-D714.

Moriyama, E. N. and J. R. Powell (1997). "Synonymous substitution rates in *Drosophila*: Mitochondrial versus nuclear genes." *Journal of Molecular Evolution* **45**(4): 378-391.

Moulton, J. K. and B. M. Wiegmann (2004). "Evolution and phylogenetic utility of CAD (rudimentary) among Mesozoic-aged Eremoneuran Diptera (Insecta)." *Molecular Phylogenetics and Evolution* **31**(1): 363-378.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck and J. L. Nieves-Aldrey (2004). "Bayesian phylogenetic analysis of combined data." *Systematic Biology* **53**(1): 47-67.

Palmén, E. (1960). "*Paratanytarsus* Arten (Dipt., Chironomidae) aus dem *ß*- nesohalinen und oligohalinen Brackwasser des Finnischen Meerbusens." *Ann. Ent. Fenn.* **26**: 280-291.

Papoucheva, E., V. Proviz, C. Lambkin, B. Goddeeris and A. Blinov (2003). "Phylogeny of the endemic Baikalian *Sergentia* (Chironomidae, Diptera)." *Molecular Phylogenetics and Evolution* **29**(1): 120-125.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Worheide and D. Baurain (2011). "Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough." *Plos Biology* **9**(3).

Pinder, L. C. V. and F. Reiss (1983). The larva of Chironominae (Diptera: Chironomidae) of the Holarctic region. Keys and diagnoses. Chironomidae of the Holarctic

region. Keys and diagnoses. Part 1. Larvae. T. Wiederholm, Entomologica scandinavica Supplement. **19:** 293-435.

Pinder, L. C. V. and F. Reiss (1986). The pupae of Chironominae (Diptera: Chironomidae) of the Holarctic region - Keys and diagnoses. Chironomidae of the Holarctic region. Keys and diagnoses. Part 2. Pupae. T. Wiederholm, Entomologica scandinavica Supplement. **28:** 299-456.

Rambaut, A. and A. J. Drummond (2008). Tracer. MCMC trace analysis tool, Institute of Evolutionary Biology, University of Edinburgh & Department of Computer Science, University of Auckland.

Raunio, J., R. Paavola and T. Muotka (2007). "Effects of emergence phenology, taxa tolerances and taxonomic resolution on the use of the Chironomid Pupal Exuvial Technique in river biomonitoring." *Freshwater Biology* **52**(1): 165-176.

Reiss, F. (1968b). "Ökologische und systematische Untersuchungen an Chironomiden (Diptera) des Bodensees. Ein Beitrag zur lakustrischen Chironomidenfauna des nördlichen Alpenvorlandes." *Archiv für Hydrobiologie* **64**(3): 247-323.

Reiss, F. (1974). "Revision des Typen-Materials einiger Tanytarsini-Arten (Chironomidae, Diptera) aus dem Museum Brüssel." *Entomologisk Tidskrift Supplement* **95**: 203-211.

Reiss, F. and E. J. Fittkau (1971). "Taxonomie und Ökologie europäisch verbreiteter *Tanytarsus*-Arten (Chironomidae, Diptera)." *Archiv für Hydrobiologie Supplement* **40**: 75-200.

Reiss, F. and L. Säwedal (1981). "Keys to males and pupae of the Palaearctic (excl. Japan) *Paratanytarsus* Thienemann & Bause, 1913, n. comb., with descriptions of three new species (Diptera: Chironomidae)." *Ent. scand. Suppl.* **15**: 73-104.

Rogers, J. and R. Wall (1980). "A mechanism for RNA splicing." *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **77**(4): 1877-1879.

Ronquist, F. and A. R. Deans (2010). Bayesian Phylogenetics and Its Influence on Insect Systematics. Annual Review of Entomology. **55:** 189-206.

Ronquist, F. and J. P. Huelsenbeck (2003). "MRBAYES 3: Bayesian phylogenetic inference under mixed models." *Bioinformatics* **19**: 1572-1574.

Rota, J. (2011). "Data partitioning in Bayesian analysis: molecular phylogenetics of metalmark moths (Lepidoptera: Choreutidae)." *Systematic Entomology* **36**(2): 317-329.

Saitou, N. and M. Nei (1986). "The number of nucleotides required to determine the branching order of 3 species, with special reference to the human-chimpanzee-gorilla divergence." *Journal of Molecular Evolution* **24**(1-2): 189-204.

Scheffler, I. (1999). Mitochondria. New York, Jonh Wiley & Sons, Inc.

Schmidt, H. A., K. Strimmer, M. Vingron and A. von Haeseler (2002). "TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing." *Bioinformatics* **18**(3): 502-504.

Silvestro, D. and I. Michalak (2011). "raxmlGUI: a graphical front-end for RAxML." *Organisms Diversity & Evolution*: 1-3.

Srivathsan, A. and R. Meier (2012). "On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature." *Cladistics* **28**(2): 190-194.

Stamatakis, A. (2006). "RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models." *Bioinformatics* **22**(21): 2688-2690.

Strimmer, K. and A. von Haeseler (1997). "Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment." *Proceedings of the National Academy of Sciences* **94**(13): 6815-6819.

Stur, E. and T. Ekrem (2006). "A revision of West Palaearctic species of the *Micropsectra atrofasciata* species group (Diptera: Chironomidae)." *Zoological Journal of the Linnean Society* **146**: 162-225.

Stur, E. and T. Ekrem (2011). "Exploring unknown life stages of Arctic Tanytarsini (Diptera: Chironomidae) with DNA barcoding." *Zootaxa*(2743): 27-39.

Sæther, O. A. and F. O. Roque (2004). "New Neotropical species of *Nandeva* (Diptera: Chironomidae), with a phylogeny of the Tanytarsini." *Tijdschrift voor Entomologie* **147**: 63-80.

Sæther, O. A. and M. Spies. (2010). "Fauna Europaea: Chironomidae." Fauna Europaea: Diptera: Nematocera Retrieved 1. May, 2012, from www.faunaeur.org.

Säwedal, L. (1982). "Taxonomy, morphology, phylogenetic relationships and distribution of *Micropsectra* Kieffer, 1909 (Diptera: Chironomidae)." *Entomologica scandinavica* **13**: 371-400.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011). "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." *Molecular Biology and Evolution* **28**(10): 2731-2739.

Thienemann, A. (1951). "*Tanytarsus*- Studien II. Die Subsectio *Paratanytarsus*. Auf Grund der nachgelassenen Papiere Friedrich Wilh. Carl Krüger's." *Archiv für Hydrobiologie Supplement* **18**: 595-632.

Thompson, J. D., D. G. Higgins and T. J. Gibson (1994). "Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Research* **22**(22): 4673-4680.

Trautwein, M. D., B. M. Wiegmann, R. Beutel, K. M. Kjer and D. K. Yeates (2012). Advances in Insect Phylogeny at the Dawn of the Postgenomic Era. Annual Review of Entomology, Vol 57. M. R. Berenbaum. Palo Alto, Annual Reviews. **57:** 449-+.

Trivinho-Strixino, S. (2010). "Two new species of *Paratanytarsus* (Diptera: Chironomidae) from southeast of Brazil." *Zootaxa*(2726): 59-67.

Velle, G., S. J. Brooks, H. J. B. Birks and E. Willassen (2005). "Chironomids as a tool for inferring Holocene climate: an assessment based on six sites in southern Scandinavia." *Quaternary Science Reviews* **24**(12-13): 1429-1462.

Ward, P. S., S. G. Brady, B. L. Fisher and T. R. Schultz (2010). "Phylogeny and Biogeography of Dolichoderine Ants: Effects of Data Partitioning and Relict Taxa on Historical Inference." *Systematic Biology* **59**(3): 342-362.

Wiegmann, B. M., M. D. Trautwein, J.-W. Kim, B. K. Cassel, M. A. Bertone, S. L. Winterton and D. K. Yeates (2009). "Single-copy nuclear genes resolve the phylogeny of the holometabolous insects." *Bmc Biology* **7**.

Wiegmann, B. M., M. D. Trautwein, I. S. Winkler, N. B. Barr, J.-W. Kim, C. Lambkin, M. A. Bertone, B. K. Cassel, et al. (2011). "Episodic radiations in the fly tree of life." *Proceedings of the National Academy of Sciences of the United States of America* **108**(14): 5690-5695.

Wiens, J. J. (2007). "Species delimitation: New approaches for discovering diversity." *Systematic Biology* **56**(6): 875-878.

Winkler, I. S., S. J. Scheffer and C. Mitter (2009). "Molecular phylogeny and systematics of leaf-mining flies (Diptera: Agromyzidae): delimitation of Phytomyza FallEn sensu lato and included species groups, with new insights on morphological and host-use evolution." *Systematic Entomology* **34**(2): 260-292.

Winterton, S. L., B. M. Wiegmann and E. I. Schlinger (2007). "Phylogeny and Bayesian divergence time estimations of small-headed flies (Diptera : Acroceridae) using multiple molecular markers." *Molecular Phylogenetics and Evolution* **43**(3): 808-832.

Xia, X. and Z. Xie (2001). "DAMBE: Software package for data analysis in molecular biology and evolution." *Journal of Heredity* **92**(4): 371-373.

Appendix

Included taxa

Table 1, Taxa included in the analysis with corresponding life stage, collection site, geographical positions and collectors.

Species	Sample ID	Life stage	Geographic origin	Latitude	Longitude	Collectors
Micropsectra borealis	NO41	M	Norway, Oppland	61.6236	8.6177	Ekrem, T.; Stur, E.
Micropsectra nana	SOD4	M	Norway, Troms	69.2604	20.3938	Dahle. S
Micropsectra roseiventris	TO84	M	Norway, Hordaland	60.4542	4.9222	Stur, E.
Paratanytarsus abiskoensis	FI112	M	Norway, Finnmark	70.3215	31.0341	Ekrem, T.; Stur, E.
Paratanytarsus austriacus	ATNA325	M	Norge, Oppland	61.9827	9.8036	Hoffstad, T.
Paratanytarsus austriacus	BJ55	M	Norge, Bjørnøya	74.4744	19.0631	Berg, OK; Finstad, A.
Paratanytarsus austriacus	BJ62	M	Norge, Bjørnøya	74.4744	19.0631	Berg, OK; Finstad, A.
Paratanytarsus austriacus	CHIR_CH477	MPL	Canada, Manitoba	58.7609	-93.9073	Ekrem, T.
Paratanytarsus austriacus	CHIR_CH554	M	Canada, Manitoba	58.7305	-93.7801	Renaud, A.
Paratanytarsus austriacus	HLC-27261	F	Canada, Manitoba	58.7800	-94.1860	?
Paratanytarsus austriacus	SOE01	M	Norge, ST	62.6903	11.8415	Aagaard, K.
Paratanytarsus austriacus	TO355	M	Germany, Bavaria	47.6290	11.1904	Ekrem, T.
Paratanytarsus austriacus	TO364	M	Canada, NWT	73.1559	-79.9740	Velle, G.
Paratanytarsus austriacus	TO365	M	Canada, NWT	73.1559	-79.9740	Velle, G.
Paratanytarsus austriacus	TO366	M	Canada, NWT	73.1559	-79.9740	Velle, G.
Paratanytarsus austriacus	TO367	M	Canada, NWT	73.1559	-79.9740	Velle, G.
Paratanytarsus bituberculatus	TO318	M	Germany, Thüringen	51.0040	10.4242	Ekrem, T; Stur, E.
Paratanytarsus dimorphis	TO435	M	Poland, Klodnolake	54.3330	18.1170	Jażdżewska, N.
Paratanytarsus dissimilis	CHIR_CH196	M	Canada, Manitoba	58.7305	-93.7801	Ekrem, T.; Stur, E.
Paratanytarsus dissimilis	TO316	M	Germany, Thüringen	51.0040	10.4242	Ekrem, T; Stur, E.
Paratanytarsus grimmii	TO18	F	Norway, Bergen	60.3852	5.3279	Halvorsen, G. A.
Paratanytarsus hyperboreus	Fi565	M	Norway, Finnmark	70.0137	23.5547	Andersen, A.
Paratanytarsus hyperboreus	AT26	M	Norway, Oppland	61.9819	9.8045	Hoffstad, T.
Paratanytarsus hyperboreus	AT48	M	Norway, Oppland	61.9819	9.8045	Ekrem, T.
Paratanytarsus hyperboreus	TO45	M	Norway, SFI	60.82749939	7.489359856	Ekrem, T.
Paratanytarsus inopertus	TO455	M	Poland, Pomerania Region	54.1930	17.9340	Giłka, W.
Paratanytarsus intricatus	TO317	M	Germany, Thüringen	51.0040	10.4242	Ekrem, T; Stur, E.
Paratanytarsus laccophilus	CHIR_CH415	M	Canada, Manitoba	58.7212	-93.7719	Ekrem, T.
Paratanytarsus laetipes	TO319	M	Germany, Thüringen	51.0040	10.4242	Ekrem, T; Stur, E.
Paratanytarsus lauterborni	Fi241	M	Norway, Finnmark	70.4427	27.3485	Ekrem, T.
Paratanytarsus natvigi	TO50	M	Norway, SFI	60.8275	7.4894	Ekrem, T.
Paratanytarsus natvigi	SOD10	M	Norway, Finnmark	69.8446	23.3735	Dahle. S
Paratanytarsus penicillatus	FI111	M	Norway, Finnmark	70.3215	31.0341	Ekrem, T.; Stur, E.
Paratanytarsus penicillatus	SOD1	M	Norway, Finnmark	69.2700	29.1184	Dahle. S
Paratanytarsus penicillatus	SOD2	M	Norway, Finnmark	69.2700	29.1184	Dahle. S
Paratanytarsus penicillatus	SOD9	M	Norway, Finnmark	69.8446	23.3735	Dahle. S
Paratanytarsus penicillatus	SOD11	M	Norway, Finnmark	69.8446	23.3735	Dahle. S
Paratanytarsus penicillatus	SOD8	M	Norway, Finnmark	69.8446	23.3735	Dahle. S
Paratanytarsus setosimanus	CHIR_CH320	M	Canada, Manitoba	58.7374	-93.8190	Ekrem, T.; Stur, E.
Paratanytarsus setosimanus	TO359	M	Norway, STY	63.6451	9.7500	Ekrem, T.; Stur, E.
Paratanytarsus sp.	HLC-27004	F	Canada, Manitoba	58.6620	-93.1890	Lankshear, J.; McGowan, J.
Paratanytarsus sp.	TO363	L	Norway, STY	63.6451	9.7500	Ekrem, T.; Stur, E.
Paratanytarsus sp.	ACTBl 9	L	Australia, NSW,	-35.3667	148.8167	Cranston, P. S.
Paratanytarsus sp.	ACTCon4	L	Australia, NSW,	-35.3667	148.8500	Cranston, P. S.
Paratanytarsus tenuis	CHIR_CH545	M	Canada, Manitoba	58.7545	-93.9979	Renaud, A.
Paratanytarsus tenuis	TO464	M	Poland, Raduńskie Dolnelake	54.3250	18.0980	Giłka, W.
Paratanytarsus tenuis	TO61	M	Norway, HOI	60.2985	6.1608	Halvorsen, G. A.
Paratanytarsus tenuis	SOD6	M	Norway, Finnmark	69.2310	29.1609	Dahle. S
Paratanytarsus tenuis	SOD7	M	Norway, Finnmark	69.2310	29.1609	Dahle. S
Tanytarsus gracilentus	SOD3	M	Norway, Troms	69.3847	20.2635	Dahle. S
Tanytarsus recurvatus	SOD5	M	Norway, Oppland	60.8632	9.5853	Dahle. S

Taxa and genes sampled

Table 2, Amplified gene fragments.

Species	ID	CAD1	CAD4	AATS1	PGD	TPI (test)	CO1*
Micropsectra borealis	NO41						
Micropsectra nana	SOD4						
Micropsectra roseiventris	TO84						
Paratanytarsus abiskoensis	FI112						
Paratanytarsus austriacus	ATNA325						
Paratanytarsus austriacus	BJ55						
Paratanytarsus austriacus	BJ62						
Paratanytarsus austriacus	CHIR_CH477						
Paratanytarsus austriacus	CHIR_CH554						
Paratanytarsus austriacus	HLC-27261						
Paratanytarsus austriacus	SOE01						
Paratanytarsus austriacus	TO355						
Paratanytarsus austriacus	TO364						
Paratanytarsus austriacus	TO365						
Paratanytarsus austriacus	TO366						
Paratanytarsus austriacus	TO367						
Paratanytarsus bituberculatus	TO318						
Paratanytarsus dimorphis	TO435						
Paratanytarsus dissimilis	CHIR_CH196						
Paratanytarsus dissimilis	TO316						
Paratanytarsus grimmii	To18						
Paratanytarsus hyperboreus	AT26						
Paratanytarsus hyperboreus	AT48						
Paratanytarsus hyperboreus	FI565						
Paratanytarsus hyperboreus	TO45						
Paratanytarsus inopertus	TO455						
Paratanytarsus intricatus	TO317						
Paratanytarsus laccophilus	CHIR_CH415						
Paratanytarsus laetipes	TO319						
Paratanytarsus lauterborni	FI241						
Paratanytarsus natvigi	SOD10						
Paratanytarsus natvigi	TO50						
Paratanytarsus penicillatus	FI111						
Paratanytarsus penicillatus	SOD1						
Paratanytarsus penicillatus	SOD11						
Paratanytarsus penicillatus	SOD2						
Paratanytarsus penicillatus	SOD8						
Paratanytarsus penicillatus	SOD9						
Paratanytarsus setosimanus	CHIR_CH320						
Paratanytarsus setosimanus	TO359						
Paratanytarsus sp.	ACTBl 9						
Paratanytarsus sp.	ACTCon4						
Paratanytarsus sp.	HLC-27004						
Paratanytarsus sp.	TO363						
Paratanytarsus tenuis	CHIR_CH545						
Paratanytarsus tenuis	SOD6						
Paratanytarsus tenuis	SOD7						
Paratanytarsus tenuis	TO464						
Paratanytarsus tenuis	TO61						
Tanytarsus gracilentus	SOD3						
Tanytarsus recurvatus	SOD5						
*CO1 only added to *P. austriacus/hyperboreus*		49	49	51	45	14	15

Sequence statistics for the total dataset

Table 3, Sequence statistics for the four gene fragments CAD1, CAD4, AATS1 and PGD in the total concatenated dataset.

Gene fragment	CAD1		CAD4		AATS1		PGD	
T	30.6		30.2		32.4		29.8	
C	15.8		14.8		16.3		18.8	
A	32.2		32.9		26.1		28.9	
G	21.4		22.1		25.1		22.5	
Length (bp) of amplified segment	861		819		423		759	
Variable sites	397	*0.46*	328	*0.40*	168	*0.40*	293	*0.39*
Parsimony informative sites	313	*0.36*	269	*0.33*	142	*0.33*	232	*0.31*

Gene trees:
CAD1

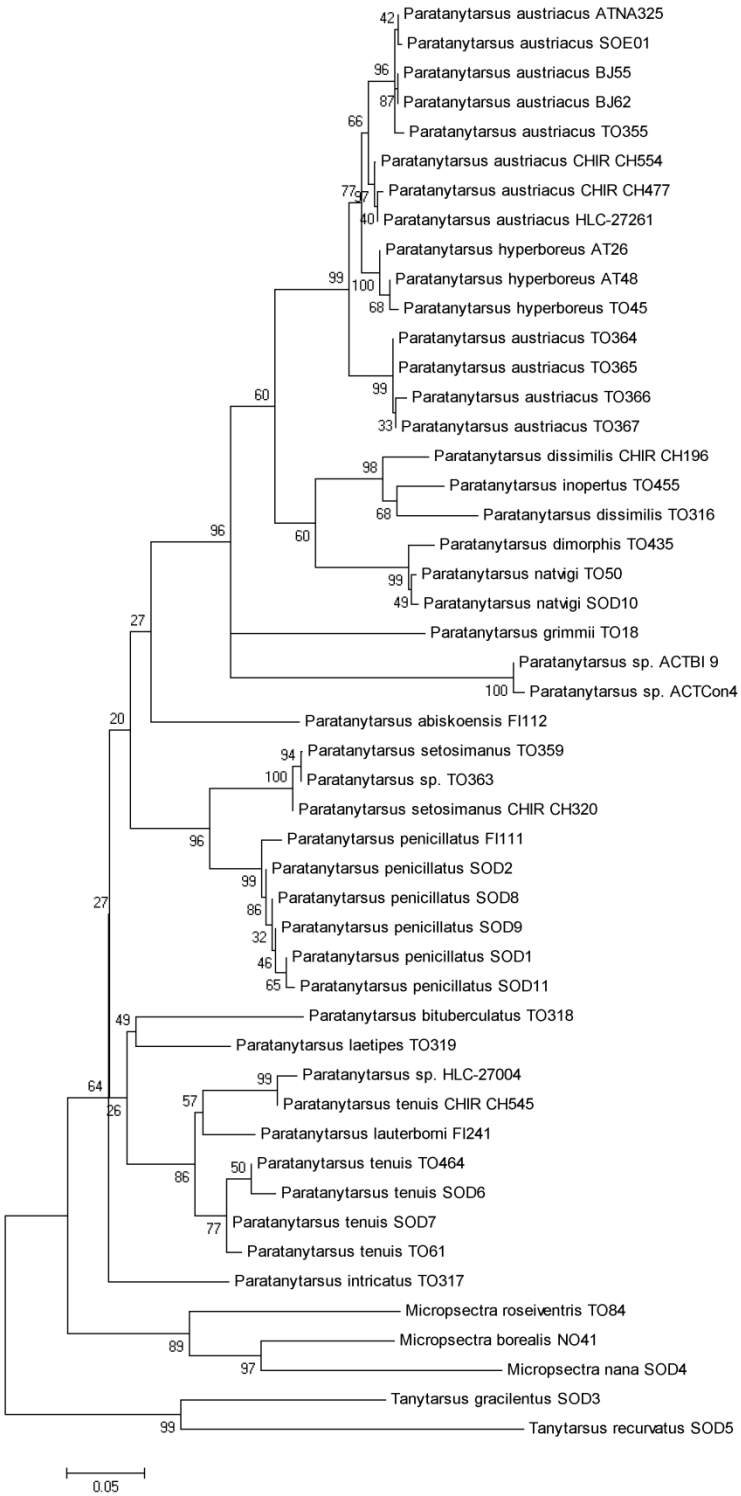


Figure 1, Bootstrap consensus tree based on CAD1, calculated in MEGA under ML and Tamura 3-parameter +G+I.

CAD4

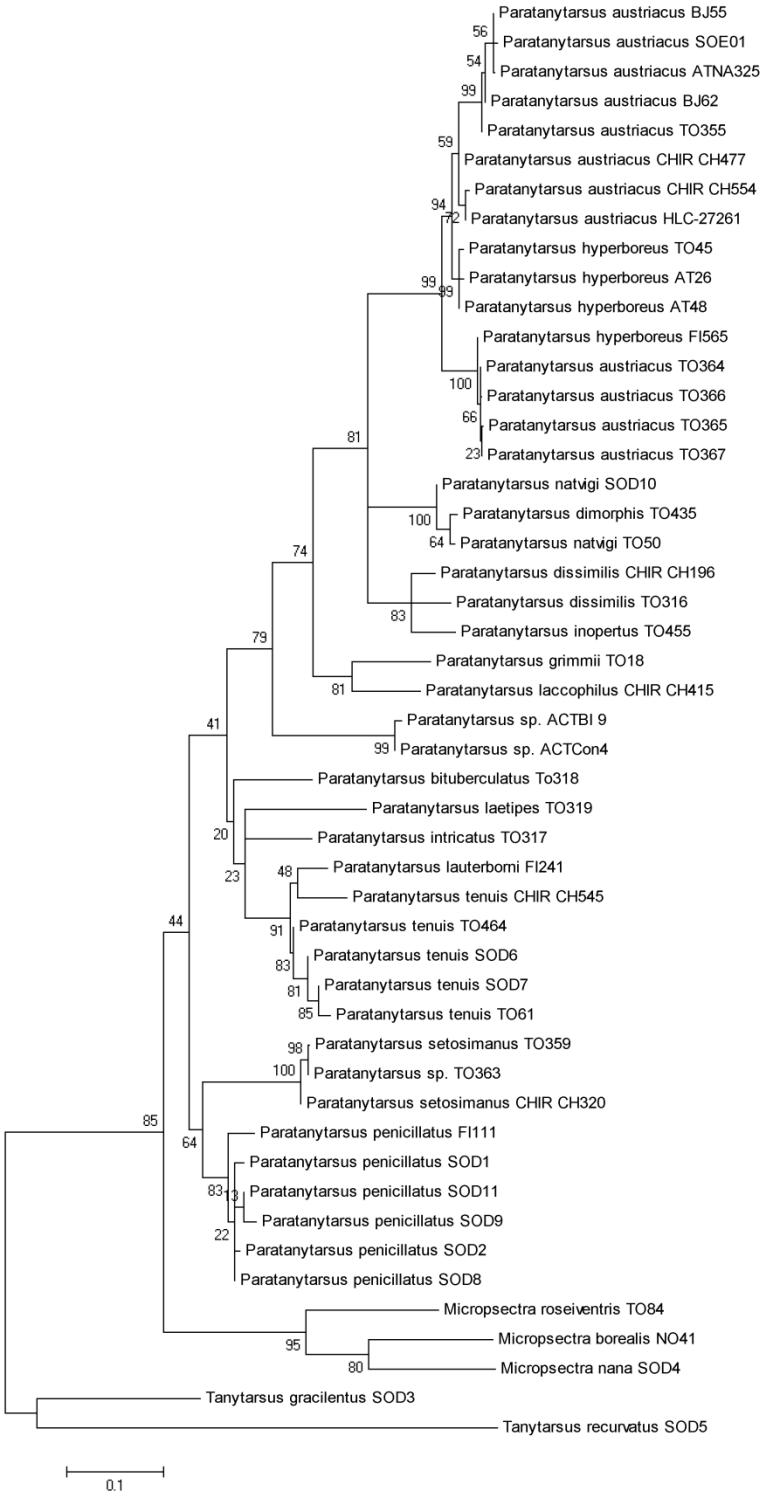


Figure 2, Bootstrap consensus tree based on CAD4, calculated in MEGA under ML and Tamura 3-parameter +G+I.

PGD

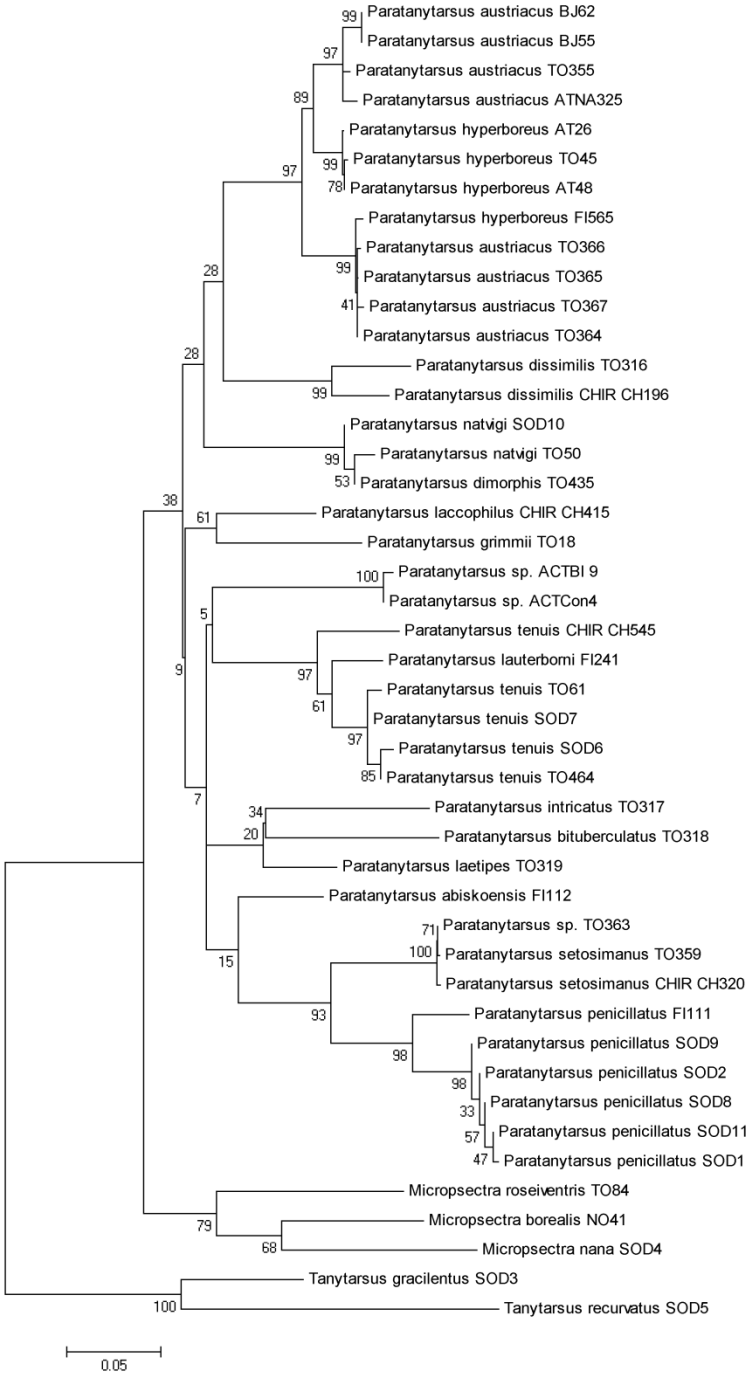


Figure 3, Bootstrap consensus tree based on PGD, calculated in MEGA under ML and Tamura 3-parameter +G+I.

AATS1

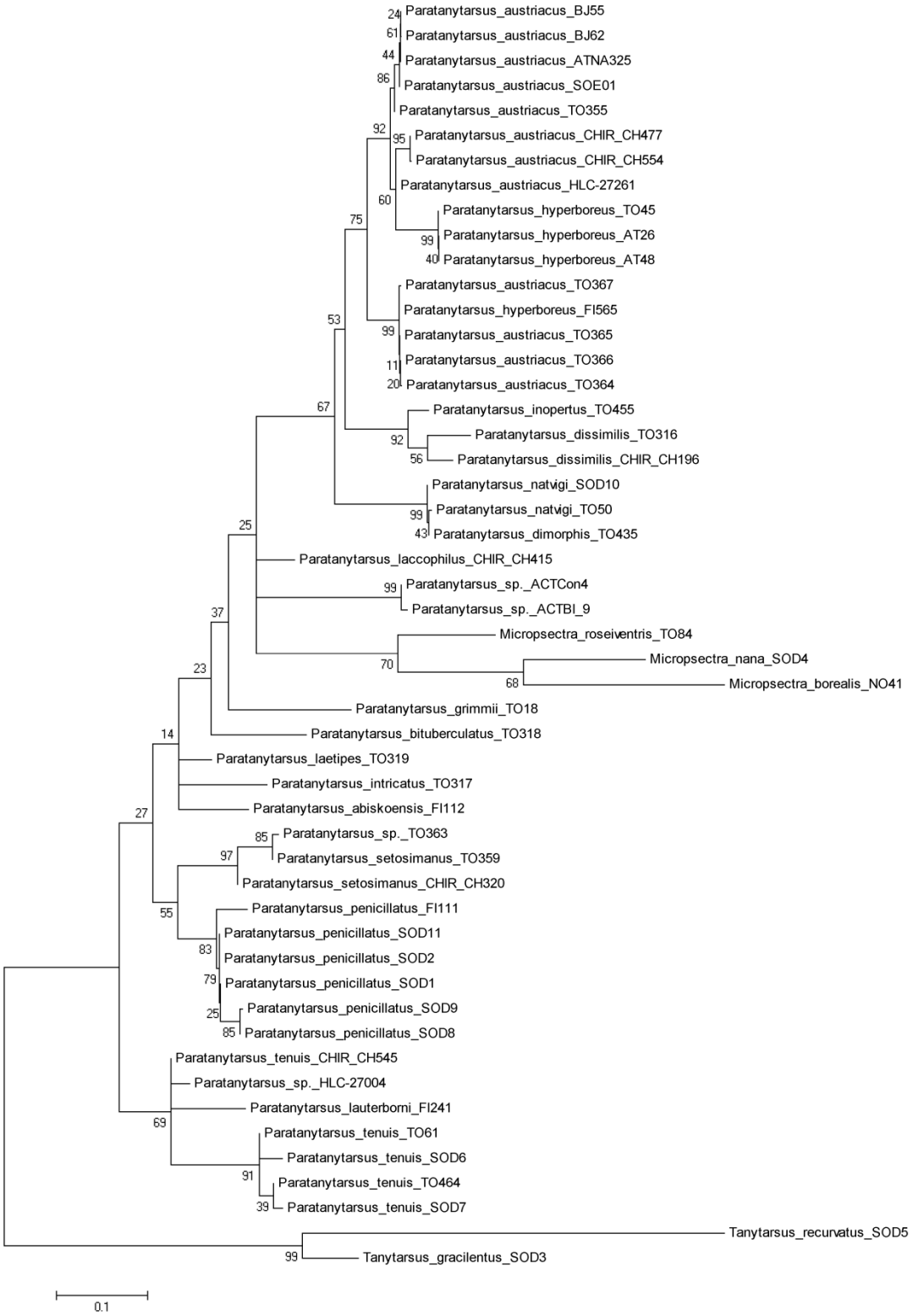


Figure 4, Bootstrap consensus tree based on AATS1, calculated in MEGA under ML and Tamura 3-parameter +G+I.

Amino acid phylogeny

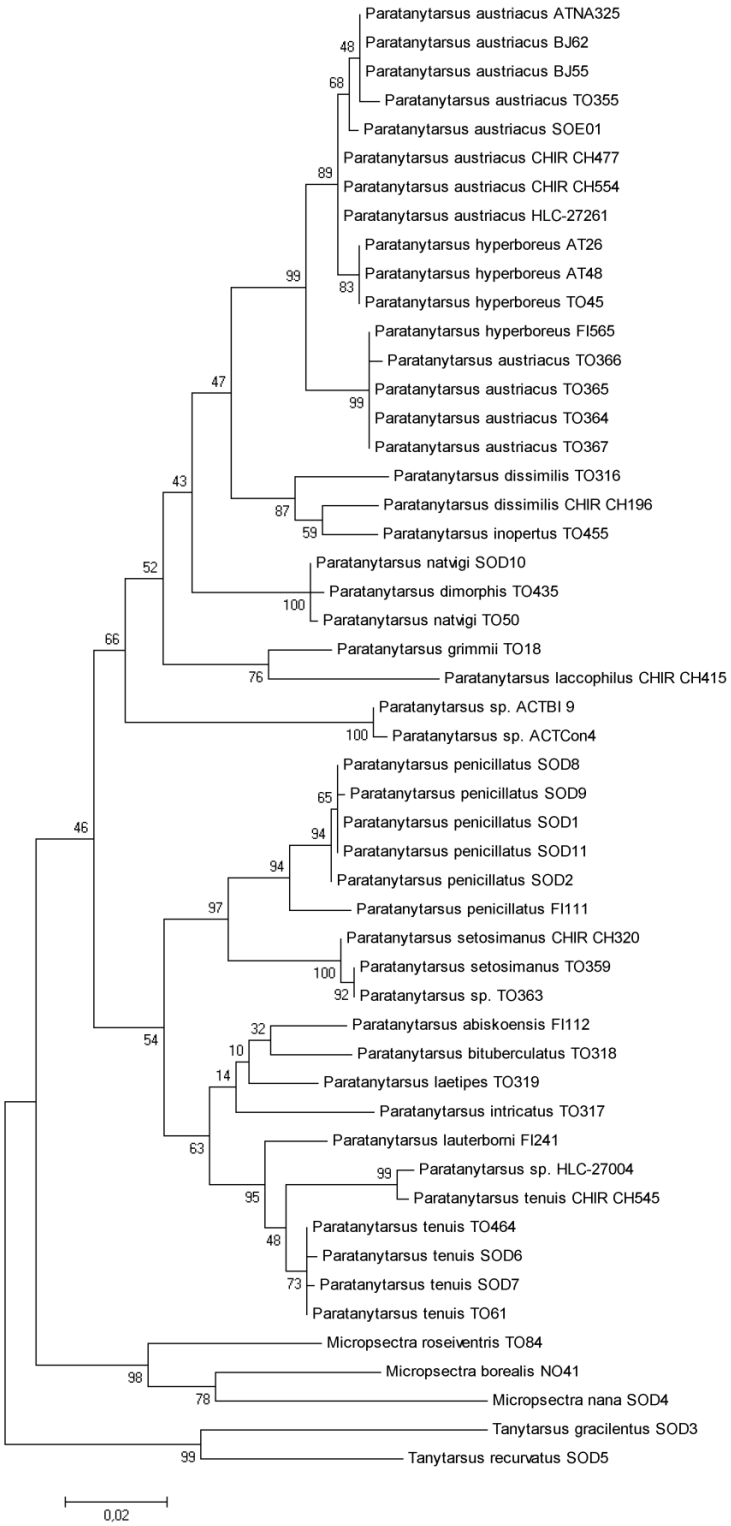


Figure 5, Result from the maximum likelihood analysis on amino acid sequences from the concatenated dataset calculated in MEGA. The Jones-Taylor-Thorton model was chosen on the basis of BIC in a maximum likelihood test performed in MEGA. Values from 1000 random bootstrap replicates are shown at branches.