Sequence and organization of the complete mitochondrial genome of the blackfly *Simulium variegatum* (Diptera: Simuliidae)

John C. Day\textsuperscript{a}, Hyun S. Gweon\textsuperscript{a} and Rory J. Post\textsuperscript{b,c}

\textsuperscript{a}Centre for Ecology and Hydrology (CEH) Wallingford, Wallingford, Oxfordshire, UK; \textsuperscript{b}School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool, UK; \textsuperscript{c}Disease Control Department, London School of Hygiene & Tropical Medicine, London, UK

**ABSTRACT**

The complete mitochondrial genome of the European blackfly, *Simulium variegatum* Meigen, 1818 was sequenced using a combined Illumina and Sanger sequencing approach. Using the known sequence of *Chironomus tepperi* Skuse, 1889 (Chironomidae) homologous NGS reads were identified and assembled. The genome is 15,367 bp in length and includes 13 protein-coding genes, 2 ribosomal RNA genes, 22 transfer RNA genes and a control region. Gene order resembles that of the ancestral dipteran gene arrangement. The base composition of the genome is A (37.6%), T (35.3%), C (15.8%) and G (11.3%). The control region between 12S rRNA and tRNA\textsubscript{Ile} is composed of 362 bp with no obvious repetitive motifs.

Blackflies belong to the family Simuliidae, a group of highly specialized aquatic insects. Adult females search for a blood meal and are notorious vectors of blood borne disease, most notably river blindness, onchocerciasis. *Simulium variegatum* (Meigen, 1818) is commonly found in small streams to large rivers throughout Ireland and the north and west of Britain. *S. variegatum* takes blood from dogs and livestock as well as man (Davies & Williams 1962). Their range extends from Algeria and Portugal in the West, throughout Europe through to Iran and the Caucasus in the East and with fourteen synonyms may be composed of a species complex (Adler & Crosskey 2015).

The *S. variegatum* specimen used in this study was collected from the River Alyn at Llandegla, Denbighshire (53.063214, -3.201966) on the 25 June 2014 as a pupa and preserved in absolute ethanol. DNA was sequenced on a HiSeq 2500 sequencer (Illumina) (San Diego, CA) and Sanger sequences. The final assembly contains 9593 Illumina sequences, 0.2% of the total dataset with an average sequence coverage of 113. The mitochondrial genome includes 13 protein-coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes and a control region between 12S rRNA and tRNA\textsubscript{Ile}, composed of 361 bp containing no apparent repetitive units. The base composition of the genome was determined to be A (37.6%), T (35.3%), C (15.8%) and G (11.3%) with a GC content of 27.0%. The annotated mitogenome, with a length of 15,367 bp, is available online at NCBI (GenBank accession number KU252587) and shares the highest sequence identity to the mitochondrial genome of *Simulium aureohirtum* (Figure 1).

We provide the first complete mitochondrial genome from a European simuliid. Little is known about the molecular biology of blackflies in comparison with other medically important Diptera. A recent study illustrated the paucity of blackfly sequences available in GenBank in comparison to other Diptera (Adler et al. 2010). However, a blackfly BAC library is now available (Crainey et al. 2010) and a current blackfly...
Figure 1. A maximum likelihood phylogenetic tree of the mitochondrial genomes *Simulium variegatum* (shown boxed) and forty six further Nematocera taxon sequences retrieved from GenBank. Sequence alignments were conducted using ClustalW and the tree constructed using PhyML with 100 bootstrap replicates using the GTR + I + G model of substitution. An alignment of 14114 characters was used for the analysis incorporating the following sequences: Aedes aegypti EU352212; Aedes albopictus KY072044; Aedes notoscriptus KM676218; Anopheles albitarsis HQ335344; Anopheles arabiensis KT382816; Anopheles atroparvus KT382817; Anopheles bellator KU551287; Anopheles christyi KT382818; Anopheles coluzzii KT382819; Anopheles craccens JX219733; Anopheles cruzii KJ701506; Anopheles culicifacies KT382820; Anopheles darlingi GQ918272; Anopheles deaneorum HQ335347; Anopheles epiroticus KT382821; Anopheles gamboa str. PEST G3 L20934; Anopheles hinesorum JX219734; Anopheles hamunculus KU551283; Anopheles kaneanus KU551288; Anopheles maculatus KT382822; Anopheles melas KT382823; Anopheles merus KT382824; Anopheles minimus KT382825; Anopheles punctulatus KT382826; Anopheles quadrimaculatus L04272; Anopheles sinensis KT218664; Anopheles stephensi KT382827; Arachnocampa flavia JN861748; Cramptonomyia spenceri JN861749; Culex pipiens pipiens HQ724614; Culex quinquefasciatus GUJ188566; Culex tritaeniorynchus KT852976; Dixella aestivalis KT878382; Haemagogus janthinomys KT372555; Nyssomyia umbratilis KP702938; Ochlerotatus vigilax KP721463; Phlebotomus chinensis KR349297; Phlebotomus papatasi KR349298; Polypedilum vanderplanki KT251040; Protoptera fitchii JN861746; Ptychoptera species JN861744; Rhopalomyia pomum GQ387649; Simulium aureohirtum KP793690; Sylvicola fenestralis JN861752; Trichocera bimaculata JN861750.
genome project is in development (Brockhouse 2015). The availability *S. variegatum* mitogenome will provide a useful resource for understanding Simuliidae evolution.

**Acknowledgements**

We would like to thank Scot E. Dowd for providing the Illumina sequencing dataset.

**Disclosure statement**

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

**Funding**

This study was supported by a NERC innovation A grant.

**References**


