The mitochondrial genome of *Gyrodactylus salaris* (Platyhelminthes: Monogenea), a pathogen of Atlantic salmon (*Salmo salar*)

T. HUYSE[1,*], L. PLAISANCE[2,3], B. L. WEBSTER[1], T. A. MO[4], T. A. BAKKE[2], L. BACHMANN[2] *and* D. T. J. LITTLEWOOD[1]

[1] *Parasitic Worms Group, Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[2] *The Natural History Museum, Department of Zoology, University of Oslo, P.O. Box 1172, Blindern, NO-0318 Oslo, Norway*
[3] *Marine Biology Research Division 0202, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0202, USA*
[4] *National Veterinary Institute, P.O. Box 8156 Dep., NO-0033 Oslo 1, Norway*

(*Received 12 September 2006; revised 25 October 2006; accepted 26 October 2006; first published online 11 December 2006*)

SUMMARY

In the present study, we describe the complete mitochondrial (mt) genome of the Atlantic salmon parasite *Gyrodactylus salaris*, the first for any monogenean species. The circular genome is 14 790 bp in size. All of the 35 genes recognized from other flatworm mitochondrial genomes were identified, and they are transcribed from the same strand. The protein-coding and ribosomal RNA (rRNA) genes share the same gene arrangement as those published previously for neodermatan mt genomes (representing cestodes and digeneans only), and the genome has an overall A+T content of 65%. Three transfer RNA (tRNA) genes overlap with other genes, whereas the secondary structure of 3 tRNA genes lack the DHU arm and 1 tRNA gene lacks the TΨC arm. Eighteen regions of non-coding DNA ranging from 4 to 112 bp in length, totalling 278 bp, were identified as well as 2 large non-coding regions (799 bp and 768 bp) that were almost identical to each other. The completion of the mt genome offers the opportunity of defining new molecular markers for studying evolutionary relationships within and among gyrodactylid species.

Key words: mitochondrial genome, *Gyrodactylus salaris*, Monopisthocotylea, *Salmo salar*, gyrodactylosis.

INTRODUCTION

Gyrodactylus salaris Malmberg, 1957 is a monogenean flatworm that is among the most serious threats to wild and farmed Atlantic salmon (*Salmo salar*) today. Ectoparasitic, viviparous and with a direct life-cycle, this species can occur in such high numbers on its hosts, and throughout entire river systems, that it can cause the collapse of fish stocks. Losses are estimated at more than 15% of the total wild salmon catch in Norway alone, with concomitant costs from the loss of fisheries, tourism and the need to survey and eradicate infected stocks, amounting to more than USD 50 million per year (Mo *et al*. 2004). Salmon stocks in 46 watercourses in Norway are threatened or have been lost (Hansen *et al*. 2003; Mo and Norheim, 2005), and other fish species such as rainbow trout (*Oncorhynchus mykiss*), brook trout (*Salvelinus fontinalis*) and Arctic charr (*S. alpinus*) readily act as suitable hosts (Bakke *et al*. 2002; Robertsen *et al*. 2006). In Europe, particularly in the UK, economically important salmon populations are at severe risk from the introduction of *G. salaris* with live fish, especially rainbow trout (Peeler and Thrush, 2004). No convincing, parasite-specific control measures are yet available. Presently, the only methods to eradicate *G. salaris* include killing all fish by the use of the biocide rotenone or by aqueous aluminium to selectively kill the parasites (Soleng *et al*. 1999, 2005; Poléo *et al*. 2004). Despite all eradication attempts, the parasite is still increasing its geographical distribution and in approximately 30% of the treated rivers, the parasite has reappeared (Mo and Norheim, 2005). Many EU states have instigated codes of practice for the movement of *G. salaris* (see Peeler *et al*. 2006), but there are few tools available to accurately track and understand the origins, spread and epidemiology of infections (Bakke *et al*. 2006).

The accurate diagnosis of infection is central to the control of *G. salaris*. Traditional methods for identification of gyrodactylids are based on morphometry of the attachment sclerites. However, more than 400 nominal species have been described within the genus *Gyrodactylus* (Harris *et al*. 2004), and additional species are continuously discovered as

* Corresponding author: Parasitic Worms Group, Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. Tel: +442 079426115. Fax: +442079425151. E-mail: T.Huyse@nhm.ac.uk

doi:10.1017/S0031182006002010 Printed in the United Kingdom

Table 1. List of PCR primer combinations for all overlapping regions

Primers	5′ to 3′
Universal*	
U12SR	TAACCGCRRMTGCTGGCACTG
U12SF	CAGTGCCAGCAKYYGCGGTTA
CytbF	GGWTAYGTWYTWCCWTGRGGWCARAT
CytbR	GCRTAWGCRAAWARRAARTAYCAYTCWGG
UNAD5F	TTRGARGCNATGCGBGCHCC
UNAD5R	GGWGCMCGCATNGCYTCYAA
UND1F	CGHAAGGGNCCNAAHAAGGT
Long PCR*	
Gsal_16SF	CCGGTGTAAGCCAGGTTGGTTC
Gsal_CytbR	GGTTAGTACCGTGGCWGCCCA
Gsal_CO1_544F	TTACTACGGATGGTGTTCGCC
Gsal_CO1_204R	GAAATACCAGACAGGTGAAGCG
*Gsal*_CytbF	TGGGCWGCCACGGTACTAACC
Gsal-MIT3	TGGCATCAATAGCCAAGCCCTTAAAGC
Sequencing primers	
16Snew	AAGTCAACATCGAGGTAGC
16S_J2	CGGTCTTAACTCAAGAGCTTCA
16STL-3′	TYACRCCGGTCTKAACTC
16STL_5′	KTRCCTTTTGYATCATG
CO1_mono5	TAATWGGTGGKTTTGGTAA
CO1_mono3	TAATGCATMGGAAAAAAACA
Platymt12S_F1	GTGCCAGDCYGCGGTTA
Gsal_COI_F1	ATCGGAGGAGTGACAGGGATAGTG
Gsal_COI_F2	CGTGTATGCTTGTATGACGACTC
Gsal_16S_R1	GCTCTTAGGGTCTTTCCGTC
Gsal_COI_F3	GAGTAGTAGGGTGGTAAACGG
Gsal_CO1_204R	GAAATACCAGACAGGTGAAGCG
Gsal_ND5_seqR	CAGGAACAGAGCATTATTAGGC
Gsal_NC2_seqR	GCCTCATCTGCCTACTTATTTG
Gsal_cox2_segF	GGAACCAGAGGCTTGTAAAATGTC
Gsal_cox2_seqF2	GGAGTTTTCGTCGGATACTG

* General and long PCR primers have also been used as sequencing primers.

new suitable host species are examined. Also, within *G. salaris*, there is significant variation in morphology. Its hook morphology closely resembles those of other related species. Discriminating among them requires high quality scanning electron micrographs and the application of sophisticated statistical tools (Shinn *et al.* 2001). However, the discrimination between *G. salaris* and *G. thymalli* Zitnan, 1960 has proved difficult, since at least 1 Norwegian *G. thymalli* population fell within the morphological variation of *G. salaris* (Shinn *et al.* 2004; unpublished observations). Thus, morphology alone does not easily lend itself to routine use in a diagnostic laboratory. To date, molecular identification methods have mainly used the internal transcribed spacers (ITS) of the nuclear ribosomal DNA (rDNA), since they display a low degree of variation within species (Zietara *et al.* 2002; Huyse *et al.* 2003; Matejusová *et al.* 2003). However, although *G. thymalli* has been described as a distinct species based on morphological, ecological and pathological grounds, no variation was detected in the ITS region when compared with *G. salaris* (see Cunningham, 1997). Limited success has been achieved with the sequencing of the intergenic spacer (IGS) of rDNA (Sterud *et al.* 2002; Cunningham *et al.* 2003; Hansen *et al.* 2006), but a suite of alternative markers is still needed for the detection of population variation, to further understand the taxonomy and biology of the parasite, and to study its transmission and dispersion in space and time.

Mitochondrial (mt) genomes offer a wealth of informative characters, useful in phylogenetic and population genetic studies (Boore *et al.* 2005). Many commercially and economically important parasites are now the focus of mitogenomic studies, as their mt genomes are now relatively easy to characterize and can be used as the basis for the design of molecular markers (Place *et al.* 2005). Thus far, partial sequences of the cytochrome *c* oxidase subunit 1 gene (*cox1*) have been used in studying the phylogeographic structure of *G. salaris* and *G. thymalli* (see Hansen *et al.* 2003, 2006; Meinila *et al.* 2004). These studies have revealed a high level of intra- and interspecific variation and could not confirm monophyly of either species. However, more variable regions are likely to be present in the other parts of the mt genome, offering higher resolution for studies on more recent evolutionary processes in *G. salaris*, such as host switching and speciation processes.

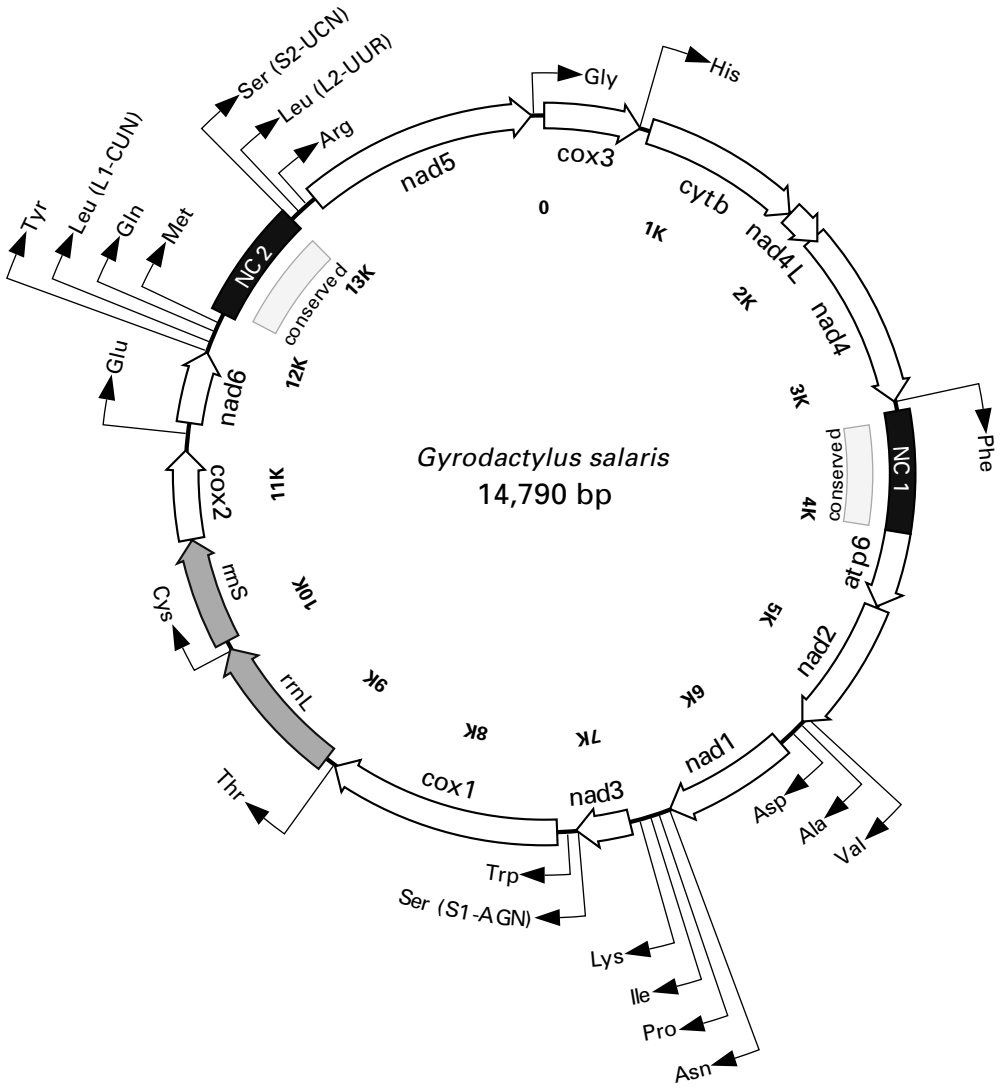

Fig. 1. Mitochondrial genome map of *Gyrodactylus salaris* from Atlantic salmon in the River Signaldalselva, North Norway. Protein-coding genes are shown as open arrows, ribosomal RNA genes as shaded arrows, and tRNAs as arrowed lines. The non-coding regions NC1 and NC2 share a 722 bp repeated element (see text for further details).

Comparing the mt genome sequences of different species will identify regions of high sequence divergence. In the present study, we sequenced and characterized the mt genome of *G. salaris*, the first for any monogenean. We describe the gene order, codon usage, tRNA features and gene boundaries, and compare these features with those found in other parasitic flatworms. This study provides the first step to defining new molecular mt markers, which will assist in studying the evolution, ecology and epidemiology of this important fish pathogen and its control.

MATERIALS AND METHODS

Sample collection and preparation

Parasite samples of *G. salaris* from wild Atlantic salmon parr were collected in the river Signaldalselva, North Norway in September 2001. This river is infected *via* migrating infected parr from the river Skibotnelva, which, in turn, has been infected by an accidental introduction of Swedish salmon parr. Hansen *et al*. (2003) sequenced the partial *cox1* of this strain (haplotype B), and found this haplotype in other Swedish rivers which drain into the Baltic Sea and are close to the Hölle hatchery, Sweden. The authors proposed that River Signaldalselva is most likely infected with *G. salaris* specimens originating from the Hölle hatchery, which is the type locality for this species. The samples were kept in 96% (v/v) ethanol at 5 °C. For genomic DNA extraction, a total of 5–10 parasites were briefly air dried to remove the ethanol, pooled in 5 μl of H_20 and digested by the addition of 5 μl of lysis solution, 1× PCR buffer (Eurogentec; 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl (pH 8·8 at 25 °C), 0·01% Tween-20), 0·45% (v/v) Tween 20, 0·45% (v/v) NP 40 and 60 μg/ml of proteinase K (Sigma). The

Table 2. Start and stop positions for individual protein-coding genes in the mitochondrial genome of *Gyrodactylus salaris* from River Signaldalselva, North Norway

(Inferred start and stop codons are indicated (see also Fig. 1).)

	Length		Codon					
Genes	bp	aa	Start	Stop	Position 5′ to 3′	Genes	Length bp	Position 5′ to 3′
Protein						RNAs		
cox3	639	213	ATG	TAA	1–639	*rrnL*	956	8933–9889
cytb	1074	358	ATG	TAG	705–1778	*rrnS*	712	9950–10661
nad4L	249	83	ATG	TAA	1784–2032	*trnH*	63	639–701
nad4	1275	425	ATG	TAA	2005–3213	*trnF*	66	3216–3281
atp6	513	171	ATG	TAA	4081–4593	*trnV*	64	5457–5521
nad2	897	299	ATG	TAA	4600–5496	*trnA*	70	5520–5589
nad1	888	296	ATG	TAA	5657–6544	*trnD*	66	5590–5656
nad3	351	117	ATG	TAG	6828–7178	*trnN*	67	6555–6621
cox1	1548	516	ATG	TAA	7311–8858	*trnP*	66	6627–6692
cox2	582	194	ATG	TAA	10662–11243	*trnI*	72	6686–6757
nad6	483	161	ATG	TAA	11430–11912	*trnK*	65	6759–6824
nad5	1551	517	ATG	TAG	13156–14706	*trnS1*	72	7177–7238
						trnW	64	7243–7306
Non-coding						*trnT*	66	8868–8933
NC 1	799				3282–4080	*trnC*	60	9890–9949
NC 2	768				12184–12951	*trnE*	72	11355–11426
						trnY	67	11916–11982
						trnL1	68	11983–12050
						trnQ	63	12051–12113
						trnM	65	12120–12184
						trnS2	73	12952–13024
						trnL2	67	13020–13086
						trnR	67	13089–13155
						trnG	68	14720–14787

sample was incubated at 65 °C for 25 min, followed by 10 min at 95 °C to inactivate the proteinase.

Conventional and long polymerase chain reaction (PCR)

Part of the cytochrome *b* gene (*cytb*) was amplified (in 25 μl reaction volumes) using universal primers *CytbF – CytbR* (see Table 1) (Boore and Brown, 2000) and puRe *Taq* Ready-to-Go PCR beads (Amersham Biosciences). The solution consisted of 20 ng of gDNA and 40 pM of each PCR primer; beads contained 1·5 U *Taq* polymerase, 10 mM Tris-HCl (pH 9·0), 50 mM KCl, 1·5 mM MgCl2, 200 μM each dNTP and stabilisers including BSA. Cycling conditions were as follows: initial denaturation for 3 min at 95 °C, followed by 40 cycles of 30 sec at 95 °C, 30 sec at 45 °C, 2 min at 72 °C, and a final extension for 10 min at 72 °C. Specific reverse and forward primers (see Table 1) were designed from the sequence of this fragment and from partial *cox1* from *G. salaris* available from GenBank (Accession no. AY258372). The mt genome was amplified by long PCR using the GeneAmp XL PCR kit (Applied Biosystems) or the Expand Long Template PCR System (Roche Applied Science) in 3 fragments employing primer pairs Gsal_CO1_544F – U12SR (2015 bp) Gsal_16SF – Gsal_CytbR (6109 bp) and Gsal_CytbF – Gsal_MIT3R (6537 bp). Additional overlapping fragments were amplified, using, for example, primer pairs U12SF – UNAD5R, UND1F –Gsal_CO1_204R). Cycling conditions were: an initial denaturation for 30 sec at 94 °C, followed by 40 cycles of 20 sec at 94 °C, 30 sec at 58–65 °C, 8 min at 64–68 °C, and final extension of 10 min at 68 °C.

Cloning and sequencing

PCR products, covering the whole genome, were individually cloned using the TOPO®XL PCR Cloning Kit (Invitrogen), following manufacturer's instructions. Ten clones from each of the cloning reactions were grown for 15 h in 3 ml volumes of Luria-Bertani (LB) medium, shaking (220 rpm) at 37 °C. Plasmid DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen). Clones were examined for inserts by digestion with the restriction endonuclease *Eco*RI (Invitrogen). Two positive clones from each of the reactions were selected for sequencing, carried out using Big Dye Chemistry (version 1.1) in a 3730 DNA Analyser (Applied Biosystems). The flanking regions of the inserts were sequenced with forward and reverse M13 primers, and sequence identity was verified using the Basic Local Alignment Search Tool (BLAST) (available at www.ncbi.nih.

```
NC1 ATGGCAGGGCTATCCAGCAAAGCCAGTTATATATTACACGTTTCAAGCCAACTTGTGGCGTGAGACTTCCCCTACTGGCGTTCGGGAATCTCACCTCTGA|100
NC2 ATGGCGATGTTATCCAGCAAAGCCAGCTATATATTACACGTTTCAAGCCAACTTGTGGCGTGAGACTTCCCCTACTGGCGTTCGGGAATCTCACCTCTGA|

NC1 TTACGTTGTGCGCTCAATAACCGTGAAACCGTTGTTGAGTAGTATCAGAATTTGACATATGTCAAAAGTTGACTCTTGTAATATATAGTACAGCCAAAAA|200
NC2 TTACGTTGTGCGCTCAATAACCGTGAAACCGTTGTTGAGTAGTATCAGAATTTGACATATGTCAAAAGTTGACTCTTGTAATATATAGTACAGCCAAAAA|

NC1 GCTCCTCGTTTATTTGTTTTGAATACAAATAAGTAGGCAGATGAGGCAAAAGTTTATCATCGTGGGAAGAAACGCAAGCTAACCCATAAGACGATTGATA|300
NC2 GCTCCTCGTTTATTTGTTTTGAATACAAATAAGTAGGCAGATGAGGCAAAAGTTTATCATCGTGGGAAGAAACGCAAGCTAACCCATAAGACGATTGATA|

NC1 ACAGCAGACACACGTAGTGGTACTATGTGTGTATGCACCGATCCGTTAACTATTGGAGAAATAGGAGTCAATGAACTTCTCCCAGAAGCACTTACCAACG|400
NC2 ACAGCAGACACACGTAGTGGTACTATGTGTGTATGCACCGATCCGTTAACTATTGGAGAAATAGGAGTCAATGAACTTCTCCCAGAAGCACTTACCAACG|

NC1 TGGGTAATGCAAAATCGCCTGGTCTTAACATATTGTTTGATATGTGAAAAAAAAAATAATCCAAAAAGTGAAATGGCGGTGTATATACATATAATATATT|500
NC2 TGGGTAATGCAAAATCGTCTGGTCTTAACATATTGTTTGATATGTGAAAAAAAAAATAATCCAAAAAGTGAAATGGCGGTGTATATACATATAATATATT|

NC1 GTCTTAATTAGGGGGACGAGGCTGATATATAAAGAGAAAATAGGGTTAATTTACAGGTAAAAGTTAAAAAAAGTTAGTATCCGTTTATCCGGGAATTTCA|600
NC2 GTCTTAATTAGGGGGACGAGGCTGATATATAAAGAGAAAATAGGGTTAATTTACAGGTAAAAGTTAAAAAAAGTTAGTATCCGTTTATCCGGGAATTTCA|

NC1 TTAAGTGAAATTACCCCTATAGAACAGCGGTACTTTCAAAAAGTAGAAAAACACTAAATTTAACTGTATCCTCAAATAGGTACCCAAAAATGCAAAATTC|700
NC2 TTAAGTGAAATTACCCCTATAGAACAGCGGTACTTTCAAAAAGTAGAAAAACACTAAATTTAACTGTATCCTCAAATAGGTACCCAAAAATGCAAAATTC|

NC1 TAAAAGATTGTTTTAGCGGTAGT
NC2 TAAAAGATTGTTTTAGCGGTAGT
```

Fig. 2. Alignment of the shared portions of non-coding regions NC1 and NC2 (see Fig. 1), illustrating sequence conservation, differences (shaded bases) and putative start codon (white on black bases) although no convincing open reading frames were detected.

gov/BLAST/). The remaining fragments were sequenced by primer walking (Table 1).

Annotation of sequences

Contiguous sequence fragments were assembled using Sequencher™ 4.5 (GeneCodes Corporation). Genome annotation was performed using MacVector® 7.2.3 (Accelrys). The sequence identity of open reading frames (ORFs) was verified using BLAST, and individual genes were aligned with published mt genomes of other flatworms to identify start and stop codons. The flatworms included: Digenea – *Schistosoma japonicum* (Accession no. NC_002544), *S. mekongi* (NC_002529), *S. mansoni* (NC_002545), *S. haematobium* (NC_008074), *S. spindale* (NC_008067), *Paragonimus westermani* (NC_002354) and *Fasciola hepatica* (NC_002546); Cestoda – *Echinococcus granulosus* (NC_008075), *E. multilocularis* (NC_000928), *Hymenolepis diminuta* (NC_002767), *Taenia crassiceps* (NC_002547), *T. asiatica* (NC_004826) and *T. solium* (NC_004022). The program tRNAscan-SE 1.21 (available at www.genetics.wustl.edu/eddy/tRNAscan-SE/) was used to identify tRNAs and their secondary structures (Lowe and Eddy, 1997). The tRNAs, which were not detected by tRNA scan-SE 1.21, were identified by searching for the conserved motif YUxxxR, where xxx denotes the anticodon, and by detecting stem and loop regions by eye. The 2

Table 3. Base composition

	A%	C%	G%	T%	AT%
Protein genes					
1st codon position	30·8	16·4	22·9	29·9	60·7
2nd codon position	19·1	14·4	20·1	46·4	65·5
3rd codon position	32·1	24·0	20·7	23·1	55·2
All positions (total)	27·3	18·3	21·2	33·1	60·4
rRNA genes	33·9	14·9	17·5	33·7	67·6
tRNA genes	32·6	13·1	18·7	34·3	66·9
All coding sites	29·1	17·2	20·5	33·2	62·3
Non-coding NC1	35·8	16·5	19·5	28·2	63·0
Non-coding NC2	36·1	16·0	19·3	28·6	64·7

Table 4. Codon usage as numbers and relative percentage values (of all codons), using translation Table 9 of GenBank (Telford *et al.* 2000)

AmAcid	Codon	Number	%	AmAcid	Codon	Number	%	AmAcid	Codon	Number	%	AmAcid	Codon	Number	%
Phe	TTT	133	3·98	Leu	CTT	60	1·8	Ile	ATT	70	2·09	Val	GTT	55	1·65
Phe	TTC	169	5·06	Leu	CTC	26	0·78	Ile	ATC	46	1·38	Val	GTC	37	1·11
Leu	TTA	150	4·49	Leu	CTA	178	5·33	Ile	ATA	214	6·4	Val	GTA	103	3·08
Leu	TTG	53	1·59	Leu	CTG	74	2·21	Met	ATG	86	2·57	Val	GTG	98	2·93
Ser	TCT	38	1·14	Pro	CCT	13	0·39	Thr	ACT	40	1·2	Ala	GCT	25	0·75
Ser	TCC	39	1·17	Pro	CCC	16	0·48	Thr	ACC	28	0·84	Ala	GCC	35	1·05
Ser	TCA	53	1·59	Pro	CCA	38	1·14	Thr	ACA	37	1·11	Ala	GCA	47	1·41
Ser	TCG	15	0·45	Pro	CCG	11	0·33	Thr	ACG	26	0·78	Ala	GCG	17	0·51
Tyr	TAT	72	2·15	His	CAT	16	0·48	Asn	AAT	50	1·5	Asp	GAT	14	0·42
Tyr	TAC	116	3·47	His	CAC	40	1·2	Asn	AAC	64	1·92	Asp	GAC	54	1·62
*	TAA	9	0·27	Gln	CAA	18	0·54	Asn	AAA	56	1·68	Glu	GAA	21	0·63
*	TAG	3	0·09	Gln	CAG	9	0·27	Lys	AAG	52	1·56	Glu	GAG	51	1·53
Cys	TGT	44	1·32	Arg	CGT	11	0·33	Ser	AGT	76	2·27	Gly	GGT	52	1·56
Cys	TGC	32	0·96	Arg	CGC	11	0·33	Ser	AGC	55	1·65	Gly	GGC	31	0·93
Trp	TGA	34	1·02	Arg	CGA	15	0·45	Ser	AGA	62	1·86	Gly	GGA	37	1·11
Trp	TGG	44	1·32	Arg	CGG	9	0·27	Ser	AGG	72	2·15	Gly	GGG	82	2·45

rRNA genes were identified by BLAST searches and boundaries were determined by the terminal ends of adjacent genes and verified by comparison with homologous genes for other flatworms available from the GenBank database. Non-coding regions were scanned for repeat elements using the program Tandem Repeats Finder (Benson, 1999).

RESULTS AND DISCUSSION

The circular mitochondrial genome of *G. salaris* is 14 790 bp long and has an overall A+T content of 62·3%. The complete annotated genome has been deposited in GenBank under Accession no. DQ988931. All the genes recognized previously in the mt genomes of other flatworms were identified in *G. salaris*. The annotated mt genome is depicted in Fig. 1, and details of gene boundaries are given in Table 2. As for other flatworms, all genes are transcribed from the same strand, and the genome lacks *atp8* and is relatively compact. Fig. 2 shows that the length of individual protein-coding genes of *G. salaris* falls within the range of that found for other flatworm taxa for which mt genome sequences are available in the GenBank. Thus, the length differences among different taxa relates mainly to non-coding regions, which, for example, can range from 1·5 to 10 kb in *S. mansoni* (see Després *et al.* 1991). The order of protein-coding and rRNA genes is consistent with that of the Neodermata, in that they follow the pattern of most digeneans (except some species of *Schistosoma*) and all cestodes for which mt genome sequence data are available (Littlewood *et al.* 2006). Only short, partial, multi-gene regions of non-neodermatan ('turbellarian') flatworms exist – the macrostomid *Microstomum lineare* (AY228756) and the polyclad *Pseudostylochus intermedius* (AB049114), and these show markedly different gene orders (*cf.* Littlewood *et al.* 2006). Of the mt genomes of the Neodermata characterized thus far, only the gene order for members of the genus *Schistosoma* appears to vary (Le *et al.* 2001).

G. salaris from the Signaldalselva river system (Troms County, North Norway) was chosen for mt genome characterization, since the mitochondrial haplotype of this strain should be the same as that of the type specimens of *G. salaris*. This particular haplotype was also found in *G. salaris* from the Vindelälv and Tornioälv river systems that drain into the Baltic Sea. These rivers are believed to have been originally infected by dumping salmon smolt in 1975 from the Hölle hatchery, Sweden, the type locality for *G. salaris* (see Johnsen *et al.* 1999; Hansen *et al.* 2003). With a known provenance, the present mt genome may now provide a reference for further mitogenomic studies, testing hypotheses regarding the ecology, epidemiology and population genetics of *G. salaris*.

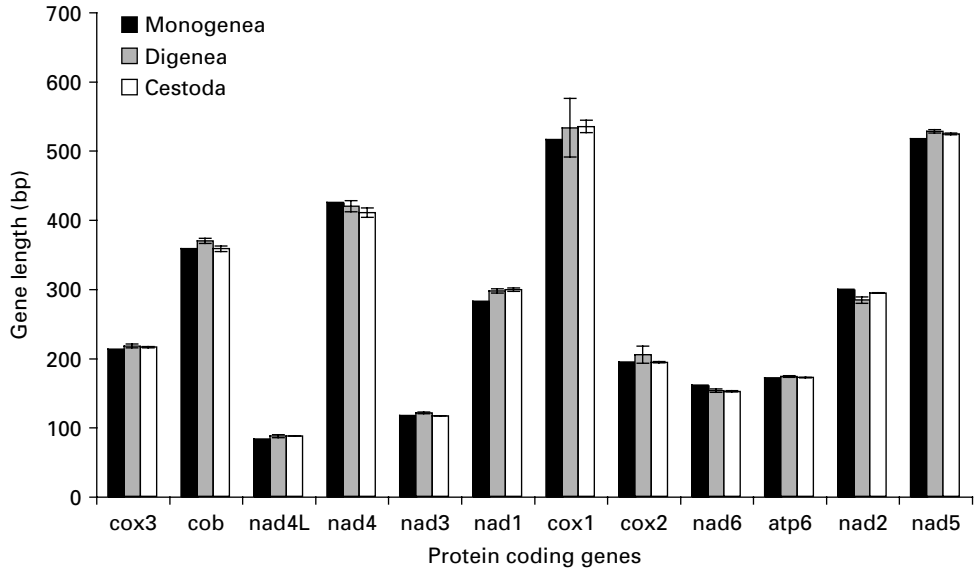

Fig. 3. Comparison of mean length of mitochondrial protein coding genes from parasitic flatworms; ± S.D. Monogenea – *G. salaris* (this study); Digenea – *Schistosoma japonicum*, *S. mekongi*, *S. mansoni*, *S. haematobium*, *S. spindale*, *Paragonimus westermani* and *Fasciola hepatica*; Cestoda – *Echinococcus granulosus*, *E. multilocularis*, *Hymenolepis diminuta*, *Taenia crassiceps*, *T. asiatica* and *T. solium* (see text for GenBank Accession numbers).

Start and stop codons

All initiation codons predicted are ATG, whereas TAG (for *cytb*, *nad3* and *nad5*) and TAA (for *cox3*, *nad4l*, *nad4*, *atp6*, *nad2*, *nad1*, *cox1*, *cox2* and *nad6*) are used as stop codons. Other invertebrate initiation codons, such as ATC and GTG, were not identified. Protein-coding genes were translated using the flatworm (rhabditophoran) mitochondrial code (see Table 9 in GenBank; Telford *et al.* 2000).

tRNA secondary structures, and overlap between adjacent genes

All of the 22 tRNA genes were identified. All but 8 were identified using the program tRNAscan; the remaining tRNA genes were identified by examining for putative anti-codon sequences and determining characteristic flanking stem and loop features. The 2 leucine tRNAs code for CTA and TTA (*trnL1* and *trnL2*, respectively), and the 2 serine tRNAs code for AGC and TCA (*trnS1* and *trnS2*, respectively). Three tRNA genes overlap with other genes, namely *trnV* (39 bp with *nad2*), *trnS1* (5 bp with *trnP*), *trnS2* (2 bp with *nad3*). Of all of the predicted secondary structures for the tRNA genes, *trnS1*, *trnS2* and *trnC* lack the DHU arm, and *trnP* appears to lack the TΨC arm.

Nucleotide composition and codon usage

The overall nucleotide composition of all coding sites is: A (29·1%), C (17·2%), G (20·5%), and T (33·2%) (see Table 3). The genome has an overall A+T content of 62·3%. The protein-coding genes display a relatively low A+T content, particularly at the third codon positions (55·2%). The highest A+T content is in the rRNA and tRNA genes (67·6 and 66·9%). These findings fall within the ranges reported for other mt genomes of flatworms (Le *et al.* 2002; Johnston 2006). The codons predicted to be most frequently used are AUA (214), CUA (187), UUC (169), UUA (150), and the least frequently used (G+C-rich) codons are CCG, CGC, CGT (11), CAG and CGG (9), not including the stop codons UAA (9) and UAG (3) (Table 4).

Non-coding regions

There are 18 non-coding regions ranging from 4 bp to 112 bp. Typical control regions are not readily identifiable within the mt genomes of flatworms. However, the origin of replication may be located in the non-coding region between *cox2* and *trnE* (positions 11244–11354), as this region folds with hairpins but without a T-rich loop (features of the control region as described by Wolstenholme (1992); nevertheless, this designation remains putative). Control regions are often targeted as a source of genetic markers, since they tend to vary considerably between species. Therefore, this region warrants characterization for additional strains and species of gyrodactylids. The 2 large non-coding regions, NC1 and NC2 (see Fig. 1) are 798 bp and 767 bp long, with an A+T content of 63 and 64·7%, respectively. NC1 is situated between *trnP* and *atp6* and NC2 within a cluster of tRNA genes (*trnW*, *trnL1*, *trnQ*, *trnM*, NC2, *trnS2*, *trnL2* and *trnR*). No repetitive regions were found in either of these non-coding regions. An alignment of NC1 and NC2 shows that almost the entire region (722 bp) is conserved (see Fig. 3), with only 6 bp (<0·1%) difference (Fig. 2).

The conserved fragment begins in frame on a recognized initiation codon, ATG. However, in spite of extensive searches, no significant ORFs (for putative genes) were detected. The stop codon TAA occurred frequently in the 3 reading frames within NC1 and NC2. Repeat regions, particularly tandemly repeated regions, can yield variable length polymorphisms through slippage events, but NC1 and NC2 are separated by several kilobases from one another, leading us to speculate as to their possible origin and function within the mt genome. One may hypothesize that the NC1 and NC2 are not of mitochondrial but of nuclear origin. However, these non-coding regions have been identified in different PCR products amplified by long PCR. None of the coding genes detected in these long PCR products displayed degeneration which is typical for nuclear copies of mt DNA. Sequences obtained from PCR products using different primer pairs were consistently the same. Furthermore, independent PCR and sequencing conducted at the Natural History Museums in London and in Oslo confirmed the results. Therefore, we conclude that a nuclear origin of the non-coding NC1 and NC2 regions can be ruled out.

In summary, the mt genome of *G. salaris*, from a pathogenic population infecting Atlantic salmon parr in northern Norway, has been sequenced and annotated. Short non-coding regions, a putative control region and 2 longer, almost identical but separated non-coding regions, offer the opportunity of discovering rapidly evolving regions of the genome suitable for the definition of strain-specific markers. The variability in these and other regions may now be explored by comparing *G. salaris* from various geographical locations with reference strains and with other species of *Gyrodactylus*. These markers can then be used to 'genotype' *G. salaris* occurring on other salmonid hosts (e.g. *Oncorhynchus mykiss*, *Salvelinus fontinalis* and *S. alpinus*) to recognize host-associated haplotypes, and to genetically characterize *G. salaris* strains with varying degrees of virulence and host-specificity to salmon (Mo, 2006; Olstad *et al*. 2006). Having available mt markers will enable tools to be developed for studying the transmission dynamics of various species of *Gyrodactylus* between/among different host species and river systems, providing crucial information for an improved understanding of the spread and epidemiology of this pathogen. Additional mt genomes from a diversity of flatworm taxa, including polyopisthocotylean monogeneans and turbellarians, will allow an assessment of the phylogenetic utility of mt genomes at deeper evolutionary levels within the phylum.

We are grateful to Dr Catherine Collins and Dr Carey Cunningham for providing starting material for this project. Dr Le Thanh Hoa provided useful primer sequences prior to their publication. We thank Julia Llewellyn-Hughes and Claire Griffin of the NHM Sequencing Unit for expert technical assistance. T.H. was funded by a Marie Curie Fellowship (Meif-ct-2004-501684). D.T.J.L. and B.L.W. were funded by the Wellcome Trust, through a Senior Research Fellowship (043965) to D.T.J.L. L.P., T.A.B. and L.B. were supported by the National Centre for Biosystematics (currently NCB 146515/420; co-funded by NHM, UiO, and the Norwegian Research Council, NRC) and the 'Wild Salmon Program' (currently NRC 145861/720). Note on authorship: T.H. and D.T.J.L. conceived the project. T.A.M. and T.B. provided parasite samples. T.H. sequenced and annotated the genome with assistance from B.L.W. and D.T.J.L. L.P. assisted with primer design. T.H. and D.T.J.L. wrote the manuscript.

REFERENCES

Bakke, T. A., Cable, J. and Harris, P. D. (2006). The biology of gyrodactylid monogeneans: the ''Russian Doll-killers''. *Advances in Parasitology* (in the Press).

Bakke, T. A., Harris, P. D. and Cable, J. (2002). Host specificity dynamics: observations on gyrodactylid monogeneans. *International Journal for Parasitology* **32**, 281–308.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–580.

Boore, J. L. and Brown, W. M. (2000). Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Molecular Biology and Evolution* **17**, 87–106.

Boore, J. L., Macey, J. R. and Medina, M. (2005). Sequencing and comparing whole mitochondrial genomes of animals. *Molecular Evolution: Producing the Biochemical Data, Part B* **395**, 311–348.

Cunningham, C. O. (1997). Species variation within the internal transcribed spacer (ITS) region of *Gyrodactylus* (Monogenea: Gyrodactylidae) ribosomal RNA genes. *Journal of Parasitology* **83**, 215–219.

Cunningham, C. O., Collins, C. M., Malmberg, G. and Mo, T. A. (2003). Analysis of ribosomal RNA intergenic spacer (IGS) sequences in species and populations of *Gyrodactylus* (Platyhelminthes: Monogenea) from salmonid fish in northern Europe. *Diseases of Aquatic Organisms* **57**, 237–246.

Després, L., Imbert-Establet, D., Combes, C., Bonhomme, F. and Monnerot, M. (1991). Isolation and polymorphism in mitochondrial DNA from *Schistosoma mansoni*. *Molecular and Biochemical Parasitology* **47**, 139–142.

Hansen, H., Bachmann, L. and Bakke, T. A. (2003). Mitochondrial DNA variation of *Gyrodactylus* spp. (Monogenea, Gyrodactylidae) populations infecting Atlantic salmon, grayling, and rainbow trout in Norway and Sweden. *International Journal for Parasitology* **33**, 1471–1478.

Hansen, H., Martinsen, L., Bakke, T. A. and Bachmann, L. (2006). The incongruence of nuclear and mitochondrial DNA variation supports conspecificity of the monogenean parasites *Gyrodactylus salaris* and *G. thymalli*. *Parasitology* **133**, 639–650.

Harris, P. D., Shinn, A. P., Cable, J. and Bakke, T. A. (2004). Nominal species of the genus *Gyrodactylus* von

Nordmann 1832 (Monogenea: Gyrodactylidae), with a list of principal host species. *Systematic Parasitology* **59**, 1–27.

Huyse, T., Audenaert, V. and Volckaert, F. A. M. (2003). Speciation and host-parasite relationships in the parasite genus *Gyrodactylus* (Monogenea, Platyhelminthes) infecting gobies of the genus *Pomatoschistus* (Gobiidae, Teleostei). *International Journal for Parasitology* **33**, 1679–1689.

Johnsen, B. O., Møkkelgjerd, P. I. and Jensen, A. J. (1999). The parasite *Gyrodactylus salaris* on salmon parr in Norwegian rivers, status report at the beginning of year 2000. (In Norwegian, English summary). *NINA Oppdargsmelding* **617**, 1–129.

Johnston, D. A. (2006). Genomes and genomics of parasitic flatworms. In *Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Physiology* (ed. Maule, A. G. and Marks, N. J.), pp. 37–80. CABI, Wallingford.

Le, T. H., Blair, D. and McManus, D. P. (2002). Mitochondrial genomes of parasitic flatworms. *Trends in Parasitology* **18**, 206–213.

Le, T. H., Humair, P. F., Blair, D., Agatsuma, T., Littlewood, D. T. J. and McManus, D. P. (2001). Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. *Molecular and Biochemical Parasitology* **117**, 61–71.

Littlewood, D. T. J., Lockyer, A. E., Webster, B. L., Johnston, D. A. and Le, T. H. (2006). The complete mitochondrial genomes of *Schistosoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. *Molecular Phylogenetics and Evolution* **39**, 452–467.

Lowe, T. M. and Eddy, S. R. (1997). tRNAscan-SE: a program for improved transfer RNA detection in genomic sequence. *Nucleic Acids Research* **25**, 955–964.

Matejusová, I., Gelnar, M., Verneau, O., Cunningham, C. O. and Littlewood, D. T. J. (2003). Molecular phylogenetic analysis of the genus *Gyrodactylus* (Platyhelminthes: Monogenea) inferred from rDNA ITS region: subgenera versus species groups. *Parasitology* **127**, 603–611.

Meinila, M., Kuusela, J., Zietara, M. S. and Lumme, J. (2004). Initial steps of speciation by geographic isolation and host switch in salmonid pathogen *Gyrodactylus salaris* (Monogenea: Gyrodactylidae). *International Journal for Parasitology* **34**, 515–526.

Mo, T. A. (2006). Chapter 2.1.14. Gyrodactylosis (*Gyrodactylus salaris*) [online manual]. World Organization for Animal Health. http://www.oie.int/eng/normes/fmanual/A_00031.htm

Mo, T. A. and Norheim, K. (2005). The surveillance and control programme for *Gyrodactylus salaris* in Atlantic salmon and rainbow trout in Norway. In *Annual Report 2004*, pp. 137–139. National Veterinary Institute.

Mo, T. A., Norheim, K. and Hellesnes, I. (2004). The surveillance and control programme for *Gyrodactylus salaris* in Atlantic salmon and rainbow trout in Norway. (In Norwegian, English summary.) *Norsk Veterinærtidsskrift* **3**, 157–163.

Olstad, K., Robertsen, G., Bachmann, L. and Bakke, T. A. (2006). Intraspecific differences in host preference among *Gyrodactylus salaris* (Monogenea) strains: an experimental approach. *Parasitology* (in the Press).

Peeler, E. J. and Thrush, M. A. (2004). Qualitative analysis of the risk of introducing *Gyrodactylus salaris* into the United Kingdom. *Diseases of Aquatic Organisms* **62**, 103–113.

Peeler, E. J., Thrush, M. A., Paisley, L. and Rodgers, C. (2006). An assessment of the risk of spreading the fish parasite *Gyrodactylus salaris* to uninfected territories in the European Union with the movement of live Atlantic salmon (*Salmo salar*) from coastal waters. *Aquaculture* **258**, 187–197.

Place, A. R., Feng, X. J., Steven, C. R., Fourcade, H. M. and Boore, J. L. (2005). Genetic markers in blue crabs (*Callinectes sapidus*) II. Complete mitochondrial genome sequence and characterization of genetic variation. *Journal of Experimental Marine Biology and Ecology* **319**, 15–27.

Poléo, A. B. S., Schjolden, J., Hansen, H., Bakke, T. A., Mo, T. A., Rosseland, B. O. and Lydersen, E. (2004). The effect of various metals on *Gyrodactylus salaris* (Platyhelminthes, Monogenea) infections in Atlantic salmon (*Salmo salar*). *Parasitology* **128**, 1–9.

Robertsen, G., Hansen, H., Bachmann, L. and Bakke, T. A. (2006). Arctic charr (*Salvelinus alpinus*) is a suitable host for *Gyrodactylus salaris* (Monogenea, Gyrodactylidae) in Norway. *Parasitology* (in the Press).

Shinn, A. P., Gibson, D. I. and Sommerville, C. (2001). Morphometric discrimination of *Gyrodactylus salaris* Malmberg (Monogenea) from species of *Gyrodactylus* parasitising British salmonids using novel parameters. *Journal of Fish Diseases* **24**, 83–97.

Shinn, A. P., Hansen, H., Olstad, K., Bachmann, L. and Bakke, T. A. (2004). The use of morphometric characters to discriminate specimens of laboratory-reared and wild populations of *Gyrodactylus salaris* and *G. thymalli* (Monogenea). *Folia Parasitologica* **51**, 239–252.

Soleng, A., Poleo, A. B. S., Alstad, N. E. W. and Bakke, T. A. (1999). Aqueous aluminium eliminates *Gyrodactylus salaris* (Platyhelminthes, Monogenea) infections in Atlantic salmon. *Parasitology* **119**, 19–25.

Soleng, A., Poleo, A. B. S. and Bakke, T. A. (2005). Toxicity of aqueous aluminium to the ectoparasitic monogenean *Gyrodactylus salaris*. *Aquaculture* **250**, 616–620.

Sterud, E., Mo, T. A., Collins, C. M. and Cunningham, C. O. (2002). The use of host specificity, pathogenicity, and molecular markers to differentiate between *Gyrodactylus salaris* Malmberg, 1957 and *G. thymalli* Zitnan, 1960 (Monogenea: Gyrodactylidae). *Parasitology* **124**, 203–213.

Telford, M. J., Herniou, E. A., Russell, R. B. and Littlewood, D. T. J. (2000). Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proceedings of the National Academy of Sciences, USA* **97**, 11359–11364.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Reviews in Cytology* **141**, 173–216.

Zietara, M. S., Huyse, T., Lumme, J. and Volckaert, F. A. (2002). Deep divergence among subgenera of *Gyrodactylus* inferred from rDNA ITS region. *Parasitology* **124**, 39–52.