Identification and characterisation of 17 polymorphic candidate genes for response to parasitic nematode (*Trichostrongylus tenuis*) infection in red grouse (*Lagopus lagopus scotica*)

Marius A. Wenzel[1*], Lucy M.I. Webster[1, 2], Steve Paterson[3] & Stuart B. Piertney[1]

August 20, 2014

[1] Institute of Biological and Environmental Sciences, University of Aberdeen, Zoology Building, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

[2] Current address: Science and Advice for Scottish Agriculture, Roddinglaw Road, Edinburgh EH12 9FJ, UK

[3] Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK

* corresponding author. email address: marius.a.wenzel.08@aberdeen.ac.uk. Phone number: +44 1224 272395

Word counts: 143 (abstract), 777 (main text), 92 (table legends)

Keywords: red grouse; Lagopus; parasite; candidate genes; adaptive genetic diversity; population genomics

Abstract

The red grouse (*Lagopus lagopus scotica*) is an economically important game bird species endemic to the upland heather moors of the British Isles, where its conservation status is "amber" due to long-term declines in breeding populations. One major driver of grouse population ecology is chronic infection by the highly prevalent, gastrointestinal parasitic nematode *Trichostrongylus tenuis*. Here, we outline the identification and characterisation of 17 candidate genes for the physiological response of red grouse to parasite infection, developed *de novo* from functional and genetic analysis of grouse transcriptomic and genomic resources. These genes capture broad physiological functions, including immune system processes, xenobiotics detoxification, oxidative balance, metabolism and cell cycle regulation. All genes were polymorphic at the landscape scale in north-east Scotland, indicating great utility for characterising the causes and consequences of spatio-temporal genetic variation in relation to parasite-mediated eco-evolutionary processes in red grouse populations.

An increasingly important aspect of conservation practice is the effective management of adaptive genetic diversity in natural populations (Ouborg et al, 2010). However, the identification of appropriate genomic regions that directly relate to traits influencing individual fitness and population viability has proven a major challenge (Allendorf et al, 2010). Here, we describe a strategy for the "top-down" identification (*sensu* Piertney and Webster, 2010) of novel polymorphic candidate genes from transcriptomic and genomic resources. Specifically, we identify 17 candidate genes for red grouse (*Lagopus lagopus scotica* Lath.) that are directly related to interactions with the highly prevalent parasitic nematode *Trichostrongylus tenuis* Mehlis (Wilson, 1983; Shaw and Moss, 1989). Chronic infection by this parasite substantially impacts grouse condition, survival and fecundity (Hudson, 1986; Watson et al, 1987; Hudson et al, 1992; Delahay et al, 1995), with negative consequences for population dynamics and long-term population viability (Hudson et al, 1998; Redpath et al, 2006; Martínez-Padilla et al, 2014).

Transcriptome libraries for caecum, spleen and liver were prepared from grouse either experimentally infected with *T. tenuis* larvae or treated with an anthelmintic (Webster et al, 2011a). Using suppression subtractive hybridisation (SSH), libraries were enriched for transcripts present in infected birds only (Webster et al, 2011a).

Clone sequences of enriched (SSH) and non-enriched (standard cDNA) libraries were used to construct a microarray for assaying differences in caecal gene transcription levels among grouse with natural parasite loads, experimental infection or anthelmintic treatment (Webster et al, 2011b). Based on gene product identity and function (BLASTX and GENEONTOLOGY; Webster et al, 2011a,b), 578 clone sequences (447 Kbp) were then used to construct a genomic capture array (Paterson *et al.*, unpublished) for identifying population-level genetic polymorphisms (SNPs) in two red grouse populations (Catterick, England and Edinglassie, Scotland) that differ in typical parasite load, and one willow grouse (*L. l. lagopus*) population from Sweden. Hybridised genomic DNA was pyrosequenced and reads were assembled to contigs. Polymorphic sites in each contig were identified (coverage $\geqq 30$ and $\geqq 6$ variant reads) and pairwise genetic differentiation (F_{ST}) among the three populations was calculated and tested for statistical significance by permutation.

Candidate contigs had to satisfy at least one of four criteria: 1) expressed in infected red grouse only (SSH libraries); 2) significantly differentially regulated ($p < 0.05$) among red grouse with different parasite loads; 3) significantly genetically differentiated ($F_{ST} > 0$; $p < 0.05$) among red grouse populations with different parasite loads (candidate for directional selection); or 4) not significantly differentiated ($F_{ST} \geq 0$; $p > 0.05$) among red grouse populations but at least weakly ($F_{ST} > 0$) among red and willow grouse (candidate for balancing selection). The functional categories of the selected candidates included immune system processes, xenobiotics detoxification, oxidative balance, metabolism and cell cycle regulation, capturing a broad physiological response to parasite infection (Table 1).

The cDNA clone sequence of each candidate contig was mapped to the chicken genome (*Gallus gallus* galGal4 assembly) using BLAT (Kent, 2002) to identify exonic genomic regions. Associated grouse genomic contigs were mapped to the identified chicken chromosome regions in GENEIOUS v5.6.3 (Drummond et al, 2012). Primers were then designed on those genomic contigs, using PRIMER3 (Rozen and Skaletsky, 2000) as implemented in GENEIOUS, so that a 120–600 bp amplicon would be at least partially exonic and include at least one polymorphic exonic site. Cross-species utility of the primers was tested using IN SILICO PCR (Hinrichs et al, 2006) on chicken (*Gallus gallus* galGal4 assembly), turkey (*Meleagris gallopavo* melGal1 assembly) and zebrafinch (*Taeniopygia guttata* taeGut1 assembly) genomes.

Levels of polymorphism were ascertained in three red grouse individuals from locations that maximise geographic variation across a landscape of grouse moors in north-east Scotland (Glenlivet 57.29 °N 3.18 °W, Mar Lodge 56.95 °N 3.66 °W and Invermark 56.89 °N 2.88 °W). PCRs were carried out in a total volume of 25 µl containing ~25 ng DNA template, 10 mM Tris-HCl, 50 mM KCl, 2.5 mM $MgCl_2$, 0.2 mM of each nucleotide, 0.5 µM of each primer and 0.625 U *Taq* DNA Polymerase (Sigma-Aldrich). PCR profiles consisted of initial denaturation at 95 ºC for 2 min, 35 cycles of denaturation at 95 ºC for 30 s, annealing at locus-specific temperatures (Table 2) for 30 s and elongation at 72 ºC for 30 s, and final elongation at 72 ºC for 5 min. In TouchDown profiles (Don et al, 1991), the annealing temperature was decreased by 0.5 ºC per cycle for the first 20 cycles (Table 2). Amplicons were purified using a QIAQUICK PCR Purification Kit (Qiagen) and Sanger-sequenced using the forward primer on an ABI 3730xl sequencer (Eurofins MWG, Ebersberg, Germany). Sequences were aligned in GENEIOUS and heterozygote sites were coded as IUPAC degenerate bases. Polymorphic sites, numbers of haplotypes, nucleotide diversity, haplotype diversity and Tajima's D (neutrality test) were then computed on reconstructed haplotypes (PHASE method) in DNASP v5 (Librado and Rozas, 2009).

Twelve genes amplified *in silico* in at least one bird model, demonstrating a degree of cross-species utility (Table 2). Polymorphism ranged from 1–13 SNPs and 2–4 haplotypes per gene (haplotype FASTA file available in electronic supplementary materials), with evidence for departure from neutrality in gene Lls_CG06 (Table 2). These genes provide a valuable resource for exploring spatio-temporal patterns of genetic variation in relation to parasite-mediated eco-evolutionary processes in red grouse populations (Table 2).

Table 1: Candidate genes for host-parasite interactions in red grouse. Location in chicken genome (chromosome) is provided alongside information on transcriptomic (library type and transcription fold change for high vs. low parasite load) and genomic capture experiments.

		Gene product information			Transcriptomics			Genetic differentiation (F_{ST})		
ID	Chr.	Name	BLASTX descriptor	Key GENEONTOLOGY terms	Library	Fold change	Clone accession	High vs. low parasite load	Red vs. willow grouse	Contig accession
Lls_CG01	1	tcb	T-cell receptor beta chain T17T-22	GO:0045087 innate immune response; GO:0006958 complement activation, classical pathway	SSH	↑	GW699322.1	–	*	KJ886553
Lls_CG02	3	cp2k4	Cytochrome P450 2K4	GO:0009636 response to toxin; GO:0016709 oxidoreductase activity	SSH	n.s.	GW703288.1	–	*	KJ886554
Lls_CG03	3	gll9	Gallinacin-9	GO:0031640 killing of cells of another organism	SSH	↓	GW702903.1	*	+	KJ886555
Lls_CG04	1	gstk1	Glutathione S-transferase kappa 1	GO:0006749 glutathione metabolic process; GO:0005777 peroxisome	SSH	n.s.	GW700181.1	***	+	KJ886556
Lls_CG05	5	capr1	Caprin-1	GO:0000932 cytoplasmic mRNA processing body; GO:0010494 stress granule	SSH	n.s.	GW702813.1	*	***	KJ886557
Lls_CG06	7	ud11	UDP-glucuronosyltransferase 1-1	GO:0017144 drug metabolic process; GO:0070069 cytochrome complex	SSH	n.s.	GW699780.1	***	***	KJ886558
Lls_CG07	1	at1a1	Sodium/potassium-transporting ATPase subunit alpha-1	GO:0007165 signal transduction; GO:0071436 sodium ion export	SSH	n.s.	GW699867.1	***	+	KJ886559
Lls_CG08	9	trfm	Melanotransferrin	GO:0006959 humoral immune response; GO:0005515 protein binding	cDNA	n.s.	GW703155.1	*	***	KJ886560
Lls_CG09	21	ccnl1	Cyclin-L1	GO:0006355 regulation of transcription, DNA-dependent	cDNA	n.s.	GW704568.1	**	+	KJ886561
Lls_CG10	1	spcs2	Signal peptidase complex subunit 2	GO:0006465 signal peptide processing	cDNA	n.s.	GW705575.1	***	+	KJ886562
Lls_CG11	2	mio	WD repeat-containing protein mio	GO:0005515 protein binding	cDNA	n.s.	GW705630.1	***	+	KJ886563
Lls_CG12	9	sumo3	Small ubiquitin-related modifier 3	GO:0045892 negative regulation of transcription, DNA-dependent	cDNA	n.s.	GW703861.1	–	+	KJ886564
Lls_CG13	1	vstm5	V-set and transmembrane domain-containing protein 5	GO:0045941 positive regulation of transcription	cDNA	↑	GW703550.1	**	+	KJ886565
Lls_CG14	7	ud11	UDP-glucuronosyltransferase 1-1	GO:0017144 drug metabolic process; GO:0070069 cytochrome complex	cDNA	↓	GW704001.1	*	***	KJ886566
Lls_CG15	18	sia7a	Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1	GO:0016266 O-glycan processing	cDNA	↑	GW706050.1	*	+	KJ886567
Lls_CG16	11	cs012	Uncharacterized protein C19orf12 homolog	GO:0031966 mitochondrial membrane	cDNA	↓	GW704603.1	**	***	KJ886568
Lls_CG17	5	coch	Cochlin	GO:0005515 protein binding	SSH	n.s.	GW699066.1	**	+	KJ886569

n.s.: not significant

–: $F_{ST} = 0.001, p > 0.05$; +: $F_{ST} > 0.001, p > 0.05$

* : $p \leqq 0.05$; ** : $p \leqq 0.01$; *** : $p \leqq 0.001$

Table 2: Characterisation of primer pairs for identified candidate genes. GC content, melting temperature T_m, annealing temperature T_a (^{TD}=TouchDown) and amplicon size are presented alongside diversity statistics (transitions Ti, transversions Tv, nucleotide diversity π, haplotypes H, haplotype diversity H_d, Tajima's D) and cross-species utility based on *in silico* PCR (S=single amplicon, M=multiple amplicons).

					Amplicon size		Diversity statistics						*in silico* amplification		
Primer name	Primer sequence (5' → 3')	GC (%)	T_m (°C)	T_a (°C)	Expected	Resolved[a]	Ti	Tv	π	H	H_d	Tajima's *D*	Chicken	Turkey	Zebrafinch
Lls_CG01_F	ACCGACTGTGGCCATCTTTTCA	50	60.2	65–55^{TD}	312	265	1	0	0.002	2	0.600	1.445	–	M	–
Lls_CG01_R	CCAGTATCACCATGGATGAATTTATGT	37	55.1												
Lls_CG02_F	AGGAGAGTTGTCACTTCTAACA	40	53.3	65–55^{TD}	191	154	2	1	0.011	3	0.733	1.386	–	–	S
Lls_CG02_R	ACAGTAAGCCCACAGGAAC	52	55.8												
Lls_CG03_F	TCTGAGACCTCACTGACCAC	55	57.2	60	158	118	1	1	0.009	3	0.800	1.032	S	S	–
Lls_CG03_R	AGGTACAAGAATTCCTCCTCAG	45	54.4												
Lls_CG04_F	ACAGATCAGATTGTTCATACTGG	39	52.5	65–55^{TD}	206	163	1	1	0.004	3	0.600	-1.132	–	–	–
Lls_CG04_R	CCTCAGCTCCAAGCCCAAAAC	57	60.5												
Lls_CG05_F	AGGGATATACAGCCTCCCAACCC	56	62	68	454	413	4	3	0.006	2	0.333	-1.390	–	–	–
Lls_CG05_R	TGCAAAGGTTTGCTAGATCC	45	53.6												
Lls_CG06_F	TGGCCGAGCATCTTTCCATCCC	59	63.8	68	336	299	7	6	0.026	2	0.600	2.262**	–	M	–
Lls_CG06_R	TGTTGGGCATCAATGGTCTTGGA	47	60.1												
Lls_CG07_F	ACTCTGGTTCTCTGTAGTATCAGCCT	46	59.3	65–55^{TD}	182	142	2	0	0.005	2	0.333	-1.132	S	–	S
Lls_CG07_R	CAGCCAAGCGTATGGCTCGT	60	62.1												
Lls_CG08_F	ACGTGTGCCAAAGTAAGCAAG	47	57	65–55^{TD}	250	214	4	1	0.012	4	0.867	0.708	–	–	–
Lls_CG08_R	AGATACCACGCCAAGGCAAA	50	58.1												
Lls_CG09_F	TTCTGTGCTTGCTGTCATGT	45	55.5	65–55^{TD}	281	240	2	0	0.003	2	0.333	-1.132	–	S	–
Lls_CG09_R	TGTGAACCTCCTTGGGCCTTC	57	61.3												
Lls_CG10_F	ATACACCCTGAAGCTGACCT	50	56.5	65–55^{TD}	226	183	3	1	0.008	4	0.800	-0.676	S	–	–
Lls_CG10_R	GCTTTCTCGCACTGCTTTCCT	52	59.4												
Lls_CG11_F	TGGGCTTTTGTTCTCTTTAGGTGT	41	57.2	65–55^{TD}	184	130	0	1	0.005	2	0.600	1.445	S	S	–
Lls_CG11_R	AGTGCACAGCGAGGAAGTGGC	61	64.5												
Lls_CG12_F	CTGGAGATGGAAGATGAAGACACT	45	56.9	65–55^{TD}	124	87	1	1	0.008	3	0.600	-1.132	–	S	–
Lls_CG12_R	GGACAGATGAGAGCGAGGTGC	61	61.9												
Lls_CG13_F	TGCCATGAGCAGCTCCATTTT	47	58.5	65–55^{TD}	378	337	1	1	0.002	2	0.333	-1.132	–	S	–
Lls_CG13_R	AGCAAAGAGCAGTGCCAACA	50	58.8												
Lls_CG14_F	GACCTCCTGAACTCTGCTTC	55	56.3	61	148	113	1	0	0.005	2	0.600	1.445	–	M	–
Lls_CG14_R	TTTGAGAAAATGAACATACCTTAGGC	34	53.5												
Lls_CG15_F	AGGAGTGGAAACGCCTGGTC	60	61.5	66	515	289[b]	6	3	0.014	4	0.867	-0.013	–	–	–
Lls_CG15_R	ACACCCAGCTCCACAAAGAGCAC	56	63.5												
Lls_CG16_F	CAGAGCTCTAAGCAGCAGGGT	57	60.6	63	211	168	0	1	0.002	2	0.333	-0.933	–	–	–
Lls_CG16_R	CAAACCCCCAACAAATGCAG	50	56.2												
Lls_CG17_F	GCAGGCCGTGCTGTTGACAC	65	64.2	65–55^{TD}	334	289	3	1	0.005	2	0.333	-1.295	–	S	S
Lls_CG17_R	AGTCTAGGAAACTTTTTCAGTGTGCT	38	56.6												

[a]: 5'-trimmed after single-end Sanger sequencing
[b]: unresolvable multiple peaks in electropherogram after particular sequence length, probably due to multiple INDEL mutations

**: $p < 0.01$

Acknowledgements

This study was funded by a BBSRC studentship (MA Wenzel) and NERC grants NE/H00775X/1 and NE/D000602/1 (SB Piertney). We are grateful to Jacob Höglund for providing willow grouse samples, Mario Röder, Keliya Bai, Marianne James, Matt Oliver, Gill Murray-Dickson, François Mougeot and Jesús Martínez-Padilla for help with fieldwork, and all grouse estate factors, owners and keepers, most particularly Alistair Mitchell, Shaila Rao, Christopher Murphy, Richard Cooke and Fred Taylor, for providing access to estate game larders.

References

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697–709

Delahay RJ, Speakman JR, Moss R (1995) The energetic consequences of parasitism: effects of a developing infection of *Trichostrongylus tenuis* (Nematoda) on red grouse (*Lagopus lagopus scoticus*) energy balance, body weight and condition. Parasitology 110(4):473–482

Don R, Cox P, Wainwright B, Baker K, Mattick J (1991) 'Touchdown'PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19(14):4008

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, S S, Thierer T, Wilson A (2012) Geneious v5.6.3. Available from http://www.geneious.com

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, et al (2006) The UCSC genome browser database: update 2006. Nucleic Acids Res 34(suppl 1):D590–D598

Hudson PJ (1986) The effect of a parasitic nematode on the breeding production of red grouse. J Anim Ecol 55(1):85–92

Hudson PJ, Dobson AP, Newborn D (1992) Do parasites make prey vulnerable to predation? Red grouse and parasites. J Anim Ecol 61(3):681–692

Hudson PJ, Dobson AP, Newborn D (1998) Prevention of population cycles by parasite removal. Int S Techn Pol Inn 282:2256–2258

Kent WJ (2002) BLAT–The BLAST-Like Alignment Tool. Genome Res 12:656–664

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Method Biochem Anal 25(11):1451–1452

Martínez-Padilla J, Redpath SM, Zeineddine M, Mougeot F (2014) Insights into population ecology from long-term studies of red grouse *Lagopus lagopus scoticus*. J Anim Ecol 83(1):85–98

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma RK, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. Trends Gen 26(4):177–187

Piertney SB, Webster LMI (2010) Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. Genetica 138:419–432

Redpath SM, Mougeot F, Leckie FM, Elston DA, Hudson PJ (2006) Testing the role of parasites in driving the cyclic population dynamics of a gamebird. Ecol Lett 9(4):410–418

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Meth Mol Biol 132:365–386

Shaw JL, Moss R (1989) Factors affecting the establishment of the caecal threadworm *Trichostrongylus tenuis* in red grouse (*Lagopus lagopus scoticus*). Parasitology 99(2):259–264

Watson H, Lee DL, Hudson PJ (1987) The effect of *Trichostrongylus tenuis* on the caecal mucosa of young, old and anthelmintic-treated wild red grouse, *Lagopus lagopus scoticus*. Parasitology 94(2):405–411

Webster LMI, Mello LV, Mougeot F, Martínez-Padilla J, Paterson S, Piertney SB (2011a) Identification of genes responding to nematode infection in red grouse. Mol Ecol Res 11(2):305–313

Webster LMI, Paterson S, Mougeot F, Martínez-Padilla J, Piertney SB (2011b) Transcriptomic response of red grouse to gastro-intestinal nematode parasites and testosterone: Implications for population dynamics. Mol Ecol 20(5):920–931

Wilson GR (1983) The prevalence of caecal threadworms (*Trichostrongylus tenuis*) in red grouse (*Lagopus lagopus scoticus*). Oecologia 58:265–268