

New Zealand Tree and Giant Wētā (Orthoptera) Transcriptomics Reveal Divergent Selection Patterns in Metabolic Loci

Victoria G. Twort^{1,2,4,*}, Richard D. Newcomb^{1,3}, and Thomas R. Buckley^{1,2}

¹School of Biological Sciences, University of Auckland, New Zealand

²Manaaki Whenua – Landcare Research, Auckland, New Zealand

³The New Zealand Institute for Plant & Food Research Ltd, Auckland, New Zealand

⁴Present address: Department of Biology, Lund University, Lund, Sweden

*Corresponding author: E-mail: vtwort@gmail.com.

Accepted: April 1, 2019

Data deposition: All raw reads have been deposited at the NCBI Sequence Read Archive (SRA) under the Bioproject PRJNA381871. Assembled transcriptomes, gene alignments, and PAML summary result files are deposited on Dryad.

Abstract

Exposure to low temperatures requires an organism to overcome physiological challenges. New Zealand wētā belonging to the genera *Hemideina* and *Deinacrida* are found across a wide range of thermal environments and therefore subject to varying selective pressures. Here we assess the selection pressures across the wētā phylogeny, with a particular emphasis on identifying genes under positive or diversifying selection. We used RNA-seq to generate transcriptomes for all 18 *Deinacrida* and *Hemideina* species. A total of 755 orthologous genes were identified using a bidirectional best-hit approach, with the resulting gene set encompassing a diverse range of functional classes. Analysis of ortholog ratios of synonymous to nonsynonymous amino acid changes found 83 genes that are under positive selection for at least one codon. A wide variety of Gene Ontology terms, enzymes, and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways are represented among these genes. In particular, enzymes involved in oxidative phosphorylation, melanin synthesis, and free-radical scavenging are represented, consistent with physiological and metabolic changes that are associated with adaptation to alpine environments. Structural alignment of the transcripts with the most codons under positive selection revealed that the majority of sites are surface residues, and therefore have the potential to influence the thermostability of the enzyme, with the exception of prophenoloxidase where two residues near the active site are under selection. These proteins provide interesting candidates for further analysis of protein evolution.

Key words: metabolic rate, molecular evolution, diversifying selection, transcriptome.

Introduction

Physiological traits, such as those associated with alpine adaptation, represent interesting candidates for studies of adaptation and molecular evolution as their function is closely linked to environmental conditions (Seebacher et al. 2010; Storz et al. 2010). Adaptation to an alpine environment requires a species to overcome a series of unique challenges associated with subzero temperatures. Such challenges include increased membrane rigidity (Overgaard et al. 2005), increased ion concentration (Kostal et al. 2007), elevated oxidative stress (Lalouette et al. 2011), and increased water loss

(Denlinger and Lee 2010). How a species responds to these challenges is dependent on whether or not they employ a freeze tolerance or freeze avoidance strategy (Voituron et al. 2002). Despite the two common approaches to dealing with these challenges, the genes identified as being cold-responsive vary considerably between species with very little overlap (Qin et al. 2005; Clark and Worland 2008; Denlinger and Lee 2010; Dunning et al. 2013, 2014; Dennis et al. 2015). Although there is high interspecific variation at the gene level, enzymes associated with glycolysis, oxidative phosphorylation, and the citric acid (TCA) cycle appear repeatedly among

the genes responding to low-temperature selective pressures, both natural and artificial (Cheviron et al. 2008; Marden 2013). Variation in these genes show strong associations with fitness affecting phenotypes, such as thermogenesis and locomotor performance (Cheviron et al. 2008, 2012; Wheat et al. 2011; Marden 2013), and influence the binding affinity, kinetics, and thermostability of enzymes (Hochachka and Somero 2002; Dalziel et al. 2009).

Colonization of habitats, such as the alpine environment, requires adaptation to different selection pressures. The ability to sequence an organisms' entire genome or transcriptome provides an opportunity to investigate large scale patterns of adaptation at the molecular level. This approach has been applied to investigate adaptation to aquatic habitats in marine mammals (Foote et al. 2015), the development of electric organs (Gallant et al. 2014), the study of avian genome evolution and adaptation (Zhang et al. 2014), and the evolution of bamboo eating in pandas (Hu et al. 2017). One particular area of interest is the relationship between patterns of molecular adaptation associated with environments such as alpine habitats. High-altitude ruminants have been shown to exhibit adaptive evolution of energy-metabolism-related genes (Qiu et al. 2012; Ge et al. 2013), as well as convergent evolution of their microbiomes for energy harvesting (Zhang et al. 2016). The adaptation of metabolism-related genes and pathways is commonly observed in genome scans for selection (Wheat et al. 2011; Dunning et al. 2013; Marden 2013).

Wētā belonging to the genera *Deinacrida* and *Hemideina* are large, flightless insects endemic to New Zealand. Their distribution covers a diverse range of habitats, including montane environments (Trewick and Morgan-Richards 2005). Phylogenetic analysis suggests that alpine adaptation has evolved multiple times among this group (Morgan-Richards and Gibbs 2001). Adaptations within species of wētā, for example, variation in color (King and Sinclair 2015) and growth rate (Minards et al. 2014; Bulgarella et al. 2015), predict positive selection at the molecular level. Variation among environmental variables experienced by different species of wētā (temperature, elevation, desiccation, UV radiation, parasites, etc.) might all be expected to result in selection on metabolic genes. In particular, metabolic rate variation has been shown among some montane and lowland *Hemideina* species and populations (Minards et al. 2014; Bulgarella et al. 2015; King and Sinclair 2015).

Tree wētā (*Hemideina*) consists of seven species distributed throughout New Zealand. Most *Hemideina* species are considered arboreal, living in the forest below the tree-line (Field and Sandlant 2001), with the exception of *Hemideina maori* which is alpine adapted and takes refuge under rocks (Gibbs 1998). In comparison, the genus *Deinacrida* comprises 11 species of giant wētā which are restricted to parts of New Zealand that are relatively free of introduced mammalian predators or in alpine environments above the mammalian predator line (Gibbs 1998). Six species occupy subalpine or alpine habitats.

Alpine habitats require an organism to meet elevated ATP costs associated with growth and development at cooler temperatures (Chown and Gaston 1999; Addo-Bediako et al. 2002; Oikawa et al. 2006). This increased demand can be met in a variety of ways, including an increase in metabolic rate (Chown and Gaston 1999; Oikawa et al. 2006), coding and transcriptional variation that influences aerobic capacity and thermal tolerance (Dalziel et al. 2005; Cheviron et al. 2012), and mutations that affect thermostability of enzymes (Fields 2001). Here, we investigate two questions using comparative transcriptome data from all tree and giant wētā species, with a specific focus on evolution in coding regions. First, we search for evidence of positive selection in genes associated with metabolic processes, ATP production, and processing of metabolic by-products, such as reactive oxygen species (ROS) across the wētā phylogeny. Second, we investigate whether across the wētā phylogeny independent lineages and internal branches show evidence for variation in selective pressures.

Materials and Methods

Sample Collection

Samples were collected at sites throughout New Zealand between 2010 and 2015 (table 1). Specimens were collected by day and night searching, and transported back to Manaaki Whenua—Landcare Research, Auckland, where they were snap frozen and stored at -80°C . Samples were collected under the following permits: CA-31615-OTH; 35637-FAU; National Permit 36236; AK-31383-FAU; 38736-FAU; NO-23262-FAU; 37619-FAU; 36296-FAU.

RNA Extraction and Transcriptome Sequencing

We used RNAase, DNase and pyrogen-free plasticware, pipette tips, and reagents for RNA extraction. Total RNA was isolated from leg muscle tissue, with the exception of *Deinacrida elegans* where an antenna was sampled and *D. fallai* where multiple tissues were extracted independently (table 1). Tissue was ground in liquid nitrogen and extracted using Trizol reagent (Invitrogen) according to the manufacturer's protocol, with the following modifications: Samples were incubated overnight at room temperature in Trizol and ethanol precipitations were incubated overnight at 4°C . The resulting products were spin-column-purified with the RNeasy mini kit (Qiagen). Prior to cDNA library preparation, verification and quantification of the extractions were carried out using an Agilent 2100 Bioanalyzer (Agilent Technologies) and a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific) or Quant-iT RiboGreen RNA Reagent (Thermo Fisher Scientific), respectively. Cleaned RNA was used for cDNA library construction, with the exception of the *D. fallai* antennae sample where we used the Trizol-extracted RNA due to sample degradation during the cleaning process.

Table 1Summary of Sample Collection Details and Tissues Sequenced from *Deinacrida* and *Hemideina* Species

Species	Code	Sex	Locality	Latitude/Longitude	Date Collected	Collected by	Tissue(s) Sequenced
<i>D. carinata</i>	WT89	Male	Pig Island, Foveaux straight	44°56'44.263"S, 168°24'34.097"E, 5 m	January 13	R. Cole	Muscle
<i>D. connectens</i>	WT68	Male	Mt Robert	41°49'55"S, 172°48'37"E, 1,300 m	April 12	K.C. Burns	Muscle
<i>D. elegans</i>	WT104	Female	Mt Sommers	43°35'S, 171°19'E	December 13	M. McDonald	Antennae
<i>D. fallai</i>	WT58	Male	Tawhiti Rahi, Poor Knights Islands	35°27'S, 174°44'E, 40 m	April 12	K. Grayson	Muscle, testis, accessory gland, midgut, antennae
<i>D. heteracantha</i>	WT74	Male	Hauturu/Little Barrier Island	36°13'S, 175°3'E, 10 m	June 12	K. Lomas, D. Seldon	Muscle
<i>D. maohoenui</i>	WT57	Male	Warrenheip, Cambridge	37°56'S, 175°35'E, 89 m	February 10	C. Watts	Muscle
<i>D. parva</i> ^a	WT263	Unknown	Mt Fyffe	42°19'S, 173°37'E	April 15	M. McDonald	Muscle
<i>D. pluvialis</i>	WT158	Female	Matukituki River, Mt Aspiring National Park	44°13'55.077"S, 168°50'2.375"E, 900 m	April 14	T. Jewell	Muscle
<i>D. rugosa</i>	WT72	Male	Stephens Island	40°40'12.789"S, 173°59'48.877"E, 220 m	May 12	L. Moran	Muscle
<i>D. talpa</i> ^a	WT250	Male	Mt Faraday, Paparoa Range	42°01.696', 171°34.785', 1,297 m	February 15	T. Buckley, W. Chinn, R. Leschen	Muscle
<i>D. tibiospina</i> ^a	WT251	Male	Mt Arthur, Kahurangi National Park	41°13'0"S, 172°40'0"E, 1,400 m	February 15	I. Millar	Muscle
<i>H. broughti</i>	WT153	Male	Denniston Plateau	41°45'57.784"S, 171°46'26.284"E, 667 m	February 14	I. Millar, A. Walker	Muscle
<i>H. crassidens</i>	WT19	Male	Karori, Wellington	41°17.02'S, 174°44.45'E, 200 m	October 11	P. Richie	Muscle
<i>H. femorata</i>	WT67	Male	Banks Peninsula, Christchurch	43°44'S, 173°0'E	March 12	R. van Heugten	Muscle
<i>H. maori</i>	WT65	Male	Rock and Pillar Range, Otago	45°22'60"S, 170°7'0"E, 1,100 m	July 13	D. Wharton	Muscle
<i>H. ricta</i>	WT107	Male	Banks Peninsula, Christchurch	43°44'S, 173°0'E	July 13	R. van Heugten	Muscle
<i>H. thoracica</i>	WT16	Male	Birkdale, Auckland	36°48'0.219"S, 174°41'26.641"E, 20 m	September 11	V. Twort, G. Twort, O. Twort	Muscle
<i>H. trewicki</i>	WT99	Male	Cape Kidnappers, Hawkes Bay	39°38'24.53"S, 177°5'21.44"E, 150 m	February 13	M. Lusk	Muscle

^aSamples for which library construction and sequencing were carried out at BGI.

Libraries were constructed at either Landcare Research or the Beijing Genomics Institute (BGI, [table 1](#)).

Libraries prepared at Manaaki Whenua—Landcare Research were constructed using the Illumina TruSeq RNA kit v2 (catalog ID: RS-122-2001) with an input of $\sim 2.5 \mu\text{g}$ of total RNA. The PCR amplification step was performed for 12 cycles. Library quality was verified using a Bioanalyzer prior to sample quantification and sequencing at New Zealand Genomics Limited on the Illumina 96 HiSeq2000 platform to generate 100-bp paired-end reads.

In the case of BGI-prepared libraries, total RNA was shipped to BGI Shenzhen for library preparation and sequencing. Messenger RNA (mRNA) was isolated with the Dynabeads mRNA Purification Kit (Invitrogen), followed by shearing with RNA fragmentation reagent (Ambion) at 72°C . SuperScriptII Reverse Transcriptase (Invitrogen) was used for cDNA synthesis with random N6 primers (IDT). Second-strand synthesis was carried out with RNase H (Invitrogen) and DNA polymerase I (New England Biolabs), followed by end-repair, adaptor ligation, and agarose gel selection (250 ± 20 bp). Products were subsequently indexed and PCR amplified to finalize library preparation. Verification of the cDNA fragment size and quantity was carried out using a Bioanalyzer and an ABI StepOnePlus Real-Time PCR machine. Sequencing was carried out on an Illumina HiSeq2000 with 150-bp paired-end reads.

Transcriptome Assembly

Raw Illumina data were quality checked with FASTQC (Andrews 2010). Sequencing reads were processed to remove reads containing ambiguous bases (N's) and homopolymer stretches using Prinseq v 0.20.4 (Schmieder and Edwards 2011). Adaptor sequences were removed using Cutadapt 1.4.1 (Martin 2011) (minimum read length 50 bp). We assembled de novo transcriptomes for each species individually (*D. fallai* tissues were combined to generate an overall transcriptome) using Trinity 2.0.6 (Grabherr et al. 2011; Haas et al. 2013), set to default parameters with a minimum contig length of 100 bp. We chose the Trinity assembler as it performs well as a de novo assembler, offering a balance between completeness, accuracy of isoform detection, and contiguity of sequences (Honaas et al. 2016). Redundancy among the resulting transcripts was reduced using CD-HIT-EST (v3.1.1 similarity threshold of 95% [Li and Godzik 2006]). Ribosomal sequences were manually removed from the assembly based on similarity to known insect ribosomal sequences (downloaded October 14, 2014, accession numbers: EU676711.1, EU676708.1, EU676684.1, EU676685.1, EU676708.1, EU676657.1, AFS14483.1, AM888956.1, and M16074.1) using a BlastN (v2.2.28 [Altschul et al. 1997]) approach (1e-10 cut-off). Each transcript was analyzed separately from this point (i.e., splice variants were not taken into account). Transcriptome assembly quality and

completeness were assessed by mapping reads back onto the final assembly using Bowtie2 v2.1.0 (Langmead and Salzberg 2012) (sensitive option) and by assessing the presence of conserved arthropod genes with BUSCO (Simão et al. 2015) (transcriptome option). Open reading frames (ORFs) were identified for each species using TransDecoder, with default parameters except for the retention of the longest ORFs provided they were larger than 120 amino acids long (Haas et al. 2013).

Ortholog Identification

We used the *D. fallai* TransDecoder results to compile a reference gene set, as this transcriptome was the most complete and encompassed a variety of tissue types. All the ORFs identified from the *D. fallai* transcriptome were annotated using a BlastX search (e-value threshold $1e-05$) against the National Centre for Biotechnology Information (NCBI) Swiss-prot database (accessed December 17, 2015). Gene Ontology (GO) terms (Gene Ontology Consortium 2015), KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Kanehisa and Goto 2000; Kanehisa et al. 2016), and Enzyme Commission codes (McDonald et al. 2009) were assigned using BLAST2GO v2.8 (Conesa et al. 2005) based on the BlastX results. ORFs were then filtered using the following criteria; complete ORF, significant BlastX hit to a known protein, BLAST similarity of $<80\%$, and an e-value $<1e-10$. We choose this BLAST criterion in order to remove those samples with high similarity to the Swiss-prot database, as high similarity to a protein from a distantly related organism resulted in uninformative wētā ortholog alignments that showed a pairwise identity of close or equal to 100%. The ortholog data set was further limited to only include those ORFs that had significant BLAST similarity to known proteins (e-value $<1e^{-3}$), because those lacking similarity may represent novel proteins or highly diverged sequences or could represent untranslated regions or genomic DNA contamination. After screening, a reference set of 17,174 *D. fallai* transcripts were used for the ortholog search.

To identify one-to-one orthologs from the *D. fallai* reference set in the remaining 17 transcriptomes, we undertook a bidirectional BLAST approach, using BlastN (e-value threshold $1e-5$). BLAST results were filtered to include only those identified in all 18 species. The resulting genes were aligned using MAFFT 7.266 (Katoh et al. 2005; Katoh and Standley 2013), using the auto option to select the most appropriate alignment strategy. Alignments were screened to ensure that at least 90% of the ORF was present in all species, and stop codons were removed. A final data set representing 755 genes was taken forward for downstream analysis.

Phylogeny Reconstruction

In order to investigate the patterns of selection across the wētā phylogeny, three methods were used to reconstruct

the phylogeny. The most appropriate model of DNA evolution was identified for each gene using the Akaike information criterion from jModelTest2 (v2.1.6) (Guindon and Gascuel 2003; Darriba et al. 2012). A maximum likelihood phylogeny was constructed independently for each gene in GARLI V2.01 (Zwickl 2006) using 25 search and 1,000 bootstrap replicates. Gene trees were used as input for species tree construction with ASTRALII v4.10.11 (Mirarab and Warnow 2015), and gene-and-site level bootstrapping was performed. For comparative purposes, we constructed a Bayesian and maximum likelihood phylogeny from the concatenated nucleotide alignment partitioned by codon. The Bayesian phylogeny was constructed using ExaBayes v1.4.1 (Aberer et al. 2014) for 1 million generations, sampled every 500. A relative burn-in of 25% was used with the following priors; exponentially distributed substitution rates, unconstrained branch lengths, empirically estimated state frequencies, exponential gamma shape parameter set to five for among site rate variation, and proportion of invariable sites uniformly distributed between 0 and 1. Each data set was analyzed in four different runs. The maximum likelihood phylogeny was generated in GARLI using a partitioned model with separate GTR+I+G model for each codon position, and 1,000 bootstrap samples were used to determine the degree of support for each node.

Inferring Selection

To detect the patterns of selection occurring across the wētā phylogeny, the ratio of nonsynonymous to synonymous amino acid substitutions (ω) was estimated for all orthologs using the CODEML package of PAML v.4.5 (Yang 2007). For all comparisons the species tree generated by ASTRAL-II was used, with an unrooted tree being generated using the Ape R package (Paradis et al. 2004). Two separate sets of analyses were implemented to assess selection patterns. To detect amino acid sites under selection ($\omega > 1$), four-site model comparisons were used (M0:M3, M1a:M2a, M7:M8, and M8a:M8). The models implemented have been extensively described elsewhere (Yang 2007). Briefly, M0 (one-ratio) allows for a single ω for all sites. M1a (nearly neutral) has two categories of sites, $\omega = 1$ and $\omega < 1$. M2a (positive selection) has the same two categories as the M1a, with the addition of a site category allowing for positive selection ($\omega > 1$). The M3 (discrete) model has three categories of site, for which the ω value can vary. M7 (“beta” neutral model) has eight site categories with eight ω taken from a discrete approximation of the beta distribution (range 0–1). M8 (“beta” plus ω) has the same eight categories of site as in the M7 model, plus an additional category with an ω that can vary from 0 to >1 . While the M8a (beta plus $\omega_s = 1$) model is similar to the M8 model, with the ω_1 being fixed at 1. Multiple models and comparisons are run to test the robustness of the patterns observed. The M0:M3 tests for rate heterogeneity between amino acid sites. The other comparisons test for positive

selection, with the M1a:M2a being a more conservative test, the M7:M8 having more power, whereas M8a:M8 combines both power and robustness. Likelihood ratio tests (LRTs) between models where ω is allowed to vary above one, and the associated null models where ω is fixed at one, allows inference of the selective pressures acting along the protein sequence (Yang 2007). Codons under positive selection were identified using the BEB (Bayes Empirical Bayes) method under the M2a model.

Branch tests were implemented to identify genes under selection along individual branches or specific lineages. Each analysis had a unique branch labeled as the foreground branch. This model allows for the labeled branch to have an ω independent of the rest of the phylogeny. *P*-values generated from the LRT tests were corrected for multiple testing using a false discovery rate of 5% with the R package QValue (Storey 2002). KEGG pathway and GO enrichment analysis were carried out in Kobas (Xie et al. 2011) by first mapping transcripts to the *D. melanogaster* reference set (*e*-value threshold $1e-5$), followed by enrichment analysis using the Fisher’s exact test, with all other genes identified in the transcriptome as the background set. Thereby, testing for a non-random distribution of terms within the genes under selection.

Secondary protein structures were modeled using the SWISS-MODEL Workspace (Waterhouse et al. 2018). A BlastP search of the RCSB Protein Data Bank (Berman et al. 2000) aligned the *D. fallax* protein sequence with the closest ortholog. The resulting structural alignment was viewed in Cn3D (Wang et al. 2000).

Results

To investigate the phylogenetic relationships and selection patterns among *Deinacrida* and *Hemideina* species, 18 de novo assembled transcriptomes were generated, primarily from muscle tissue. Illumina sequencing produced a total of 194 Gb (5–38 Gb of data per species) (table 2). De novo assembly of transcriptomes from each species, followed by the removal of redundant and rRNA sequences, resulted in an average of 223,301 contigs per species. On average, the assemblies have an N50 of 851, and a total assembly length of 111 Mb. Assembly statistics for each species are given in table 2 and supplementary table 1, Supplementary Material online. Alignment of the raw reads against the final assemblies revealed that on average 82.5% of reads mapped back to transcripts, whereas 7.6% mapped to removed rRNA transcripts (table 1 and supplementary table 1, Supplementary Material online). The levels of rRNA contamination are within the expected range (Abernathy and Overturf 2016).

The quality and completeness of each transcript set were assessed using BUSCO (Simão et al. 2015). On average 1,234 of the 2,675 proteins (46%) were present (range of 41–84%, supplementary table 1, Supplementary Material online).

Table 2Summary Statistics of Transcriptome Sequencing, Assembly, and Assessment from *Deinacrida* and *Hemideina* Species

Species	Raw Data (Gb)	No. Cleaned Pairs	No. Transcripts	Total Assembly Length (bp)	BUSCO Results (Complete (Duplicated): Fragmented: Missing)	Reads Mapping to the Final Assembly (%)
<i>D. carinata</i>	5.794	53,028,680	158,184	85,686,758	62(15): 15: 11	87.78
<i>D. connectens</i>	5.021	45,723,934	143,006	70,585,929	50(10): 13: 35	87.64
<i>D. elegans</i>	3.912	35,840,262	154,541	75,027,101	52(10): 14: 33	86.68
<i>D. fallai</i> ^a	38.025	315,605,868	491,052	238,830,923	83(32): 6.5: 9.7	86.02
<i>D. heteracantha</i>	8.662	79,034,252	181,474	95,857,769	65(15): 10: 23	82.06
<i>D. mahoenui</i>	6.527	59,526,220	136,415	63,537,626	41(7.3): 15: 43	70.03
<i>D. parva</i>	8.838	58,759,734	276,182	124,660,635	69(21): 12: 18	85.69
<i>D. pluvialis</i>	20.541	185,352,010	273,395	153,195,209	84(29): 4.7: 10	88.95
<i>D. rugosa</i>	7.938	72,763,410	162,088	89,502,720	67(15): 10: 22	80.17
<i>D. talpa</i>	10.245	68,125,774	292,982	136,051,565	77(25): 9.6: 13	85.11
<i>D. tibiospina</i>	9.202	61,191,748	285,966	135,419,927	76(25): 8.6: 14	82.86
<i>H. broughi</i>	19.766	176,252,720	283,853	151,790,647	79(26): 6.9: 13	81.59
<i>H. crassidens</i>	5.621	52,140,542	206,762	96,930,441	46(10): 16: 37	73.97
<i>H. femorata</i>	6.101	56,290,568	171,783	90,810,936	63(14): 10: 26	85.27
<i>H. maori</i>	6.118	56,420,004	157,489	82,514,748	59(13): 11: 28	86.68
<i>H. ricta</i>	20.283	184,487,074	296,632	162,318,013	84(28): 4.5: 11	88.95
<i>H. thoracica</i>	5.888	53,491,706	212,966	97,372,708	46(11): 17: 35	65.02
<i>H. trewicki</i>	5.55	50,694,098	134,650	65,137,927	44(8.1): 14: 40	80.33

NOTE.—Data cleaning was carried out to remove reads with ambiguous bases, homopolymer regions and adapters. Mapping back to the transcriptome was carried out with Bowtie2 sensitive option. Completeness was assessed with BUSCO.

^a*Deinacrida fallai* transcriptome was constructed by combining reads from all six tissue libraries together.

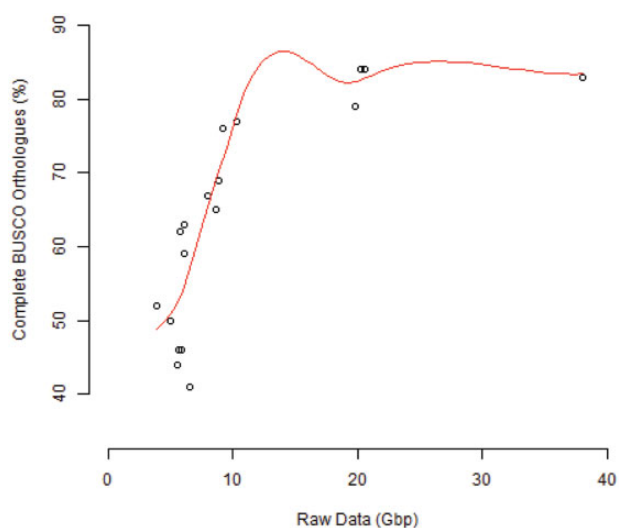


FIG. 1.—The effect of increasing data on the completeness of *Hemideina* and *Deinacrida* transcriptomes, as assessed by the presence of complete single copy arthropod orthologs (BUSCO).

Generally speaking, the variation in the completeness of each transcriptome correlated with the amount of data available for the assembly. The plateau observed when comparing the completeness of the 20-Gb assemblies with the combined *D. fallai* assembly (38 Gb of data, fig. 1) is due to diminishing returns, whereby increasing data do not yield additional or more complete transcripts (Honaas et al. 2016). The BUSCO

analysis demonstrates that despite the fact that the transcriptomes were taken from a single tissue, they still represent a useful resource for further research of wētā and downstream evolutionary analyses.

Orthologous Gene Identification

The *D. fallai* combined transcriptome was used as a basis for the generation of the reference ortholog data set. This was chosen as it contained a total of 38 Gb of raw data and had one of the highest complete sets of BUSCO orthologs (83%, [supplementary table 1, Supplementary Material](#) online). ORF identification with TransDecoder followed by filtering, resulted in a *D. fallai* reference set of 17,174 ORFs. A bidirectional best-hit approach extracted >4,000 conserved, orthologous genes from all 18 species. After filtering the results to remove sequences with <90% of the ORF, 755 genes remained for downstream analysis. The majority of orthologs were assigned to at least one GO term, with the exception of two sequences. The distribution of transcripts across all three categories (Biological Process, Molecular Function, and Cellular Compartment) was similar ([supplementary fig. 1, Supplementary Material](#) online). Of the transcripts assigned a molecular function term, the majority were assigned to binding, followed by catalytic activity and transporter activity. Cellular, single-organism and metabolic processes were the three categories with the most transcript assigned to the biological processes category. Roughly, 26% (202 transcripts) were assigned Enzyme Commission numbers, and

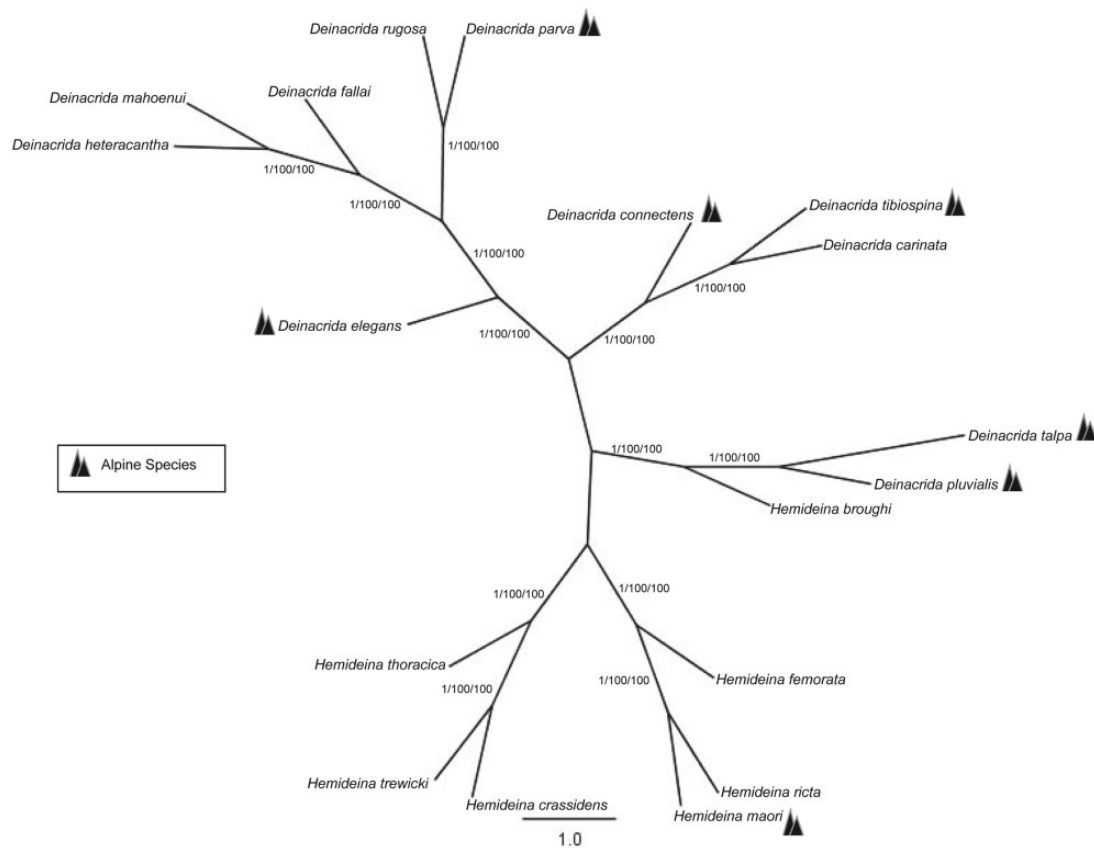


Fig. 2.—Consensus phylogeny constructed using 755 orthologs obtained from transcriptome sequences from *Hemideina* and *Deinacrida* species. Support values are Bayesian posterior probabilities/maximum likelihood bootstrap values/astral local posterior probabilities. Branch lengths are given in coalescent units, as calculated by Astral. Triangles indicate species whose range includes an alpine habitat.

Table 3

Summary of the Number of Genes Identified as Having Amino Acids under Positive Selection for Various CODEML Site Model Comparisons in PAML

Model	Number of Genes
M0:M3	473
M1a:M2a	83
M7:M8	105
M8a:M8	159

encompassed all six classes. The most prominent class identified was hydrolases, followed by transferases with 86 and 56 sequences identified, respectively ([supplementary fig. 2, Supplementary Material](#) online).

Phylogenetic Reconstruction

Three separate phylogenetic analyses were undertaken; one species tree reconstruction and two concatenated analyses. All nodes were identically reconstructed across analyses ([fig. 2](#)). Nodal posterior probabilities of 1.0 (Bayesian analysis), likelihood bootstrap values of 100%, and Astral support values of 100% were estimated for all nodes. The alpine lineages

do not form a monophyletic group with respect to lowland lineages and this is consistent with previous analyses ([Morgan-Richards and Gibbs 2001](#)).

Selection Patterns among Amino Acid Sites

Site-based likelihood models were used to determine whether positive selection events have occurred within genes associated with metabolic processes across the wētā phylogeny. We found the majority of orthologs to be evolving under purifying selection when a single ω value is calculated across the entire protein-coding sequence using the one-ratio (M0) model (mean $\omega = 0.21$, range 0.001–3.23, [supplementary fig. 3, Supplementary Material](#) online). Two sets of orthologs have an $\omega > 1$ indicating that the protein as a whole is under positive selection. BLAST homology identifies these ortholog sets as a DNA directed RNA Polymerase ii subunit and a cytochrome c-type heme lyase. Rate heterogeneity among amino acid sites was tested using the M0:M3 comparison. Despite the majority of genes being under overall purifying selection, significant among-site variation was observed for 473 of the 755 orthologs, with the M3 model, which permits three ω values, providing a significantly better fit to the data than the

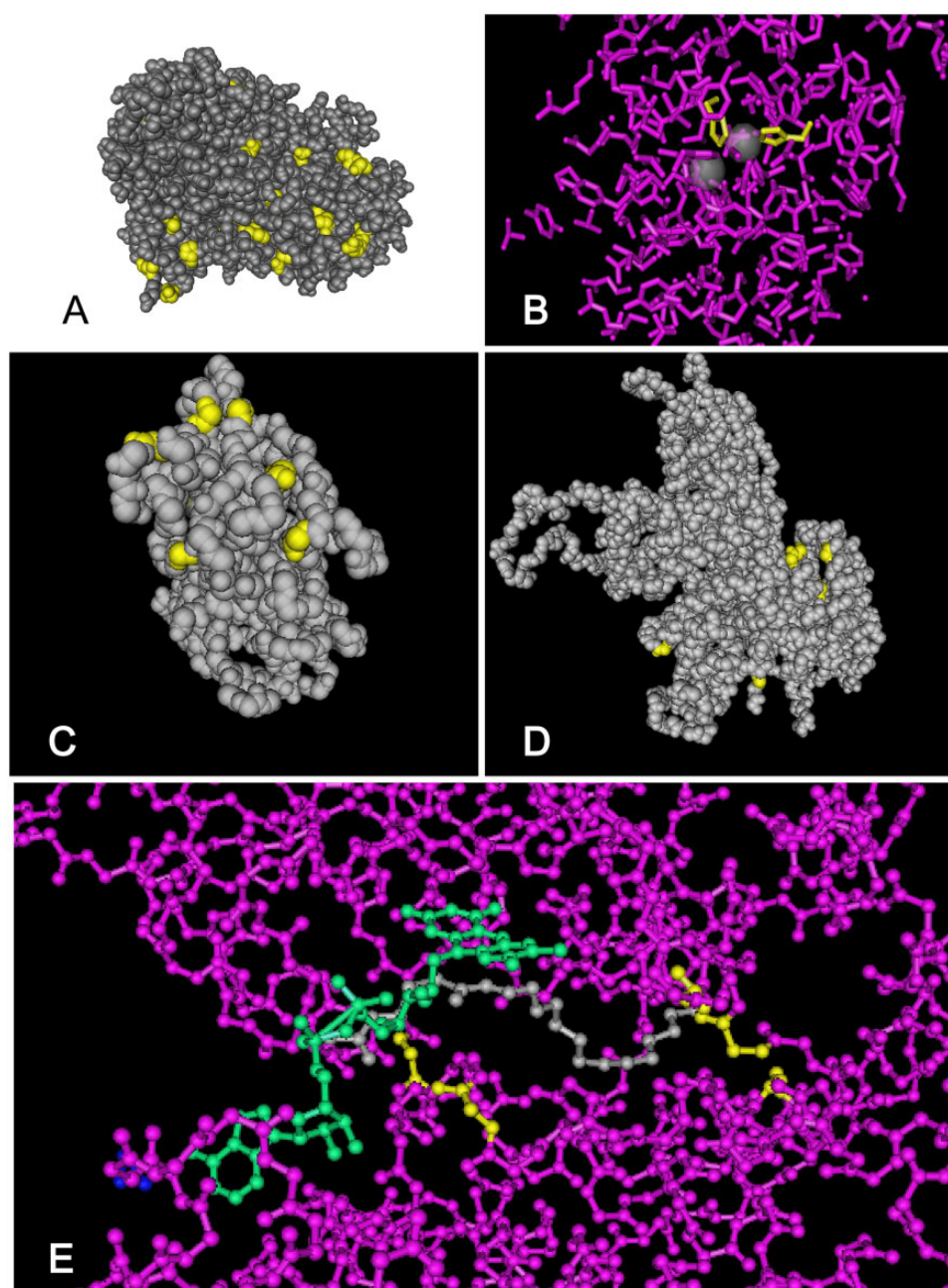


FIG. 3.—The location of sites identified as under positive selection ($\omega > 1$) using Bayes Empirical Bayes analysis under the M2a model in PAML. The structural models of the three genes are used to highlight the location in the 3D protein structure. (A) Prophenoloxidase sites under selection (shown in yellow), (B) Visual representation of prophenoloxidase active site showing the amino acids under selection (yellow), (C) superoxide dismutase positively selected amino acid residues (yellow), (D) very-long-chain-acyl-CoA dehydrogenase 3D representation showing surface residues, and (E) acyl-CoA dehydrogenase ligand (green) binding pocket sites under selection (yellow).

M0 model (corrected $P < 0.05$, 4 degrees of freedom). This, however, is not a test of positive selection. Three different model comparisons allow for the identification of positive selection among codons, each varying in their power and robustness. The M1a:M2a comparison is considered the most stringent test for positive selection, and compares the M1a null model, which has two values of ω ($\omega < 1$, $\omega = 1$), and

M2a, which has an additional category of ω that is > 1 . The M7:M8 has more power than the M1a:M2a, whereas the M8a:M8 comparison has more power and robustness. The numbers of genes identified in each model are given in [table 3](#). The M1a:M2a comparison results were used for further downstream analyses, as the evidence for selection on the genes identified is considered especially robust, whereby a

Table 4

Summary of PAML Site Test Results for the Top Ten Genes under Selection When Ranked by Number of Sites under Positive Selection

Contig	Annotation	Protein Length (aa)	K (M0)	ω (M0)	M1a:M2a	BEB Sites (M2)	RCSB Ortholog ID
3,861	Prophenoloxidase	727	2.54	0.43	82.85**	35	3HHS
255	Collagen alpha-2 chain	1778	3.56	0.29	119.46**	22	—
3,721	Very long-chain-specific acyl-mitochondrial	635	3.19	0.34	156.29**	16	2UXW
777	CD63 antigen	239	2.61	0.94	124.92**	13	—
3,921	Cytoplasmic dynein 1 intermediate chain-like isoform x16	664	3.48	0.14	59.89**	11	—
4,174	Glutathione S-transferase omega-1	242	2.57	1.01	35.74**	10	—
2,771	Superoxide dismutase	199	2.77	1.07	35.36**	9	2E47
268	Nucleoporin nup54	560	2.72	0.24	71.99**	8	—
1,338	CD63 antigen-like	231	2.72	0.87	106.75**	8	—
1,060	Microsomal glutathione S-transferase 3-like	143	3.21	0.98	28.73**	7	—

NOTE.—K, ratio of transitions to transversions; ω (M0), ratio of nonsynonymous to synonymous substitutions under the M0 model; M1a:M2a, likelihood ratio test result for the nested model comparison between the M1a and M2a model; aa, amino acid; BEB Sites (M2), number of amino acid detected under positive selection under a Bayes Empirical Bayes model.

** $P > 99\%$.

total of 83 gene sets are identified as having codons under positive selection. The number of sites identified as statistically significant under the M2a model ranges from 0 to 39 for each gene.

The orthologous sets identified from the M1a:M2a selection test encompass a diverse range of GO terms, KEGG pathways, and Enzyme Commission classes. To test for the enrichment of molecular functions, a GO enrichment test was undertaken within KOBAS. A total of 47 GO categories were enriched within the 83 orthologs ([supplementary table 2, Supplementary Material](#) online). Seven molecular function GO categories were enriched within the data set and included the terms copper iron binding (GO:0005507, ontology: Molecular Function), superoxide dismutase activity (GO:0004784, ontology: Molecular Function), and aspartic-type endopeptidase activity (GO:0004190, ontology: Molecular Function). A total of 17 biological process terms were also enriched, and these included acetyl-CoA metabolic processes (GO:0006084, ontology: Biological Process), removal of superoxide radicals (GO:0019430, ontology: Biological Process), and copper ion transport (GO:0006825, ontology: Biological Process). Five of the six Enzyme Commission classes are also represented within the data set ([supplementary fig. 4, Supplementary Material](#) online). The most prominent class of enzyme identified as under selection is hydrolases followed closely by oxidoreductases. The enzymes are distributed throughout a total of 42 KEGG pathways ([supplementary table 3, Supplementary Material](#) online). No pathways were identified as being enriched with the Fisher's exact test using KEGG pathway assignment within KOBAS. The number of enzymes present in each pathway ranged from 1 to 4. The pathways with the most enzymes identified as being under selection were purine metabolism (Pathway ID map00230), oxidative

phosphorylation (Pathway ID map00190), glutathione metabolism (Pathway ID map00480), and the biosynthesis of antibiotics (Pathway ID map01130).

The top ten transcripts with the most amino acid sites under positive selection are shown in [table 4](#). Of these transcripts, six have been identified as enzymes. Three of the six enzymes had homologous 3D protein structures in the RCSB Protein Data Bank, with enough overall similarity to model the 3D structure. Structural alignment of the wētā protein with its nearest homolog allows for the determination of where in the 3D structure the amino acid sites under selection occur. The transcript identified as the inactive precursor form of the phenoloxidase enzyme has 35 codons under positive selection, 25 of which are located on the surface of the protein and two that occur near the enzyme active site ([fig. 3A and B](#)). The majority of the sites in these alignments were on the surface ([fig. 3C and D](#)). In addition, the transcript identified as very-long-chain-acyl-CoA dehydrogenase (c131411_g2_i1|m.460001) has five sites in the ligand binding site under positive selection ([fig. 3E](#)).

Selection Patterns across the Wētā Phylogeny

To test for differences in selection patterns across the phylogeny, branch tests were carried out by independently labeling each branch of the phylogeny (33 branches in all). Of the transcripts tested between 0 and 307 had foreground ratios that were significantly different from background ratios along at least one branch, indicating a difference in selection ([supplementary fig. 5, Supplementary Material](#) online). A comparison of foreground ω values, revealed strong patterns of purifying selection ($\omega < 1$), with only a single gene, annotated as annexin, having an ω of 1.85 in a single internal branch.

Discussion

Phylogenetic Relationships among New Zealand Wētā

The species tree generated in this study agrees with previous studies based on morphological, allozymes, and cytogenetic characters (Morgan-Richards and Gibbs 2001; Trewick and Morgan-Richards 2004, 2005). The genus *Hemideina* includes two main sister clades, one consisting of northern species (*H. crassidens*, *H. thoracica*, and *H. trewicki*), and the second consisting of southern species (*H. femorata*, *H. ricta*, and *H. maori*). The grouping of *H. broughi* with *D. talpa* and *D. pluvialis* is consistent with previous reports (Morgan-Richards and Gibbs 2001; Trewick and Morgan-Richards 2004). We did recover a sister group relationship between *D. mahoenui* and *D. heteracantha*, whereas previous studies recovered a sister group relationship between *D. fallai* and *D. heteracantha* (Gibbs 2001; Morgan-Richards and Gibbs 2001; Trewick and Morgan-Richards 2004). Alpine wētā do not form a monophyletic group within the phylogeny, supporting previous findings that speciation into the montane environment has occurred multiple times in wētā, rather than being considered the retention of an ancestral state (Morgan-Richards and Gibbs 2001).

Positive Selection and the Evolution of Metabolic Associated Genes

Adaptations among wētā species, such as differences in desiccation resistance (King and Sinclair 2015), combined with differences in environmental conditions, predict positive selection acting on metabolic associated genes. Most orthologs had an $\omega < 1$, indicative of overall purifying selection. This is a common occurrence with proteins generally tending to be under purifying selection as a whole, with adaptation occurring on a subset of proteins and within proteins at selected sites/regions (Li 1997; Yang et al. 2000). This pattern was also observed among *H. thoracica* and *H. crassidens* male reproductive associated proteins (Twort et al. 2017).

The 83 genes in which we found evidence of positive selection at at least one amino acid site encompass a wide range of GO terms and KEGG pathways. Of the genes identified, 36% were annotated as enzymes. This finding is consistent with the observation that metabolic enzymes are often inferred to have higher than expected levels of functionally important variation (Marden 2013), with enzymes involved in glycolysis and the TCA cycle repeatedly being identified as responding to natural selection (Cheviron et al. 2008, 2012; Wheat et al. 2011; Dunning et al. 2013; Marden 2013). Enzymes involved in both of these pathways were among the genes under positive selection in wētā. The nonneutral variation seen in metabolic enzymes is thought to be a consequence of their moonlighting functions, whereby these enzymes also play roles outside of metabolism (Kim and Dang 2005; Fu et al. 2011; Marden 2013). Furthermore, a

variety of metabolism-related pathways had enzymes under selection. These pathways include glycolysis, oxidative phosphorylation, and the pentose phosphate pathway.

Of particular interest are the enzymes identified in the oxidative phosphorylation pathway. Coding and transcriptional variation among these genes have been observed in a variety of taxa, and are hypothesized to translate into adaptive differences in aerobic capacity and thermal tolerance (Mishmar et al. 2003; Ruiz-Pesini et al. 2004; Dalziel et al. 2005; Cheviron et al. 2008; Scott et al. 2011; Zhang and Broughton 2015). The site selection test results presented here point to selection acting on metabolic-associated processes such as ATP production and the processing of free-radicals. Adaptation to an alpine environment requires an organism to be able to deal with stress-induced responses, such as the production of ROS; therefore, the enzymes and their associated selection patterns represent interesting candidates for further analysis, including functional tests to investigate what protein changes affect enzyme function. It should be noted, however, that associations between sites under selection and protein function may not correlate (Runck et al. 2010). This in part is due to sites representing false positives of the method used (Yokoyama et al. 2008). Alternatively, the function being observed in vitro may not be the function which is being selected for (Yang and dos Reis 2011), as the secondary roles of moonlighting enzymes may be the target of selection rather than the trait of interest (Storz and Zera 2011).

The orthologs showing the strongest signal for positive selection, in terms of the number of sites identified, were further investigated. Existing homologous 3D protein structures were used to determine where in the protein sites under selection are situated. Out of the six enzymes, three had 3D structures. The majority of amino acid sites under positive selection were surface residues. The protein showing the largest number of codons under positive selection was annotated as prophenoloxidase, which is the inactive form of phenoloxidase and had 35 codons identified as under positive selection. Phenoloxidase plays a key role in melanogenesis, where it catalyzes the oxidation of phenols to quinones that polymerize to form melanin (Soderhall and Cerenius 1998). In addition to pigmentation, melanin plays a variety of roles within invertebrates including desiccation resistance thermoregulation, pathogen resistance, and immune response (Reeson et al. 1998; Barnes and Siva-Jothy 2000; Wilson et al. 2001, 2002). Within wētā the *H. maori* melanic color morph has higher desiccation resistance than the nonmelanic morph (King and Sinclair 2015), and has been shown to be associated with differences in immunity potential (Robb et al. 2003). Therefore, the high levels of variation observed within this enzyme may play a role in the ability of a species to be able to adapt to an alpine environment, and the metabolic cost associated with adaptation. The majority of sites identified under the M2a BEB analysis occur at the surface of the

enzyme, with the exception of sites 398 and 434. Surface modifications can influence thermostability (Fields 2001) and catalytic performance of enzymes (Fields and Somero 1998), and are key in maintaining the adaptive potential of a protein (Nilsson et al. 2011; Schlessinger et al. 2011). We acknowledge that multiple copies of prophenoloxidase have been found in a variety of insects (González-Santoyo and Córdoba-Aguilar 2012), with both the number of evolutionary conservation varying across taxa (Waterhouse et al. 2007). Therefore, the possibility exists that the high rate of positive selection we observe among *Deinacrida* and *Hemideina* may be a function of a diversification of prophenoloxidase among these taxa or may be caused by multiple copies in the alignment. Investigation of the gene tree for prophenoloxidase disagrees with the species tree only in the placement of *D. mahoenui*, albeit with low support values (data not shown). Screening of the *D. mahoenui* transcriptome found no additional copies of the gene, and removal of this sequence from the alignment did not affect the significance of the PAML results. However, fewer sites had significant BEB values, albeit with larger ω values, with the number dropping to 11 codons; however, the two sites near the active site are still significant (data not shown). This result suggests that the orthologs in our alignment most likely represent true homologs rather than paralogs.

The production of ROS is a byproduct of metabolism and has been attributed to exposure to stress. Superoxide dismutase is a key component of the insect antioxidant enzyme system, converting ROS to oxygen and hydrogen peroxide and has been linked to the cold hardiness of a species (Fridovich 1975; Storey KB and Storey JM 2012; Gretscher et al. 2016). A superoxide dismutase gene was identified in the top ten genes under selection. All of the sites identified as being under positive selection are surface residues and potentially influence the catalytic performance of the enzyme (Fields and Somero 1998). Within the Antarctic midge (*Belgica antarctica*), superoxide dismutase plays a role in the protection of high levels of ROS in intense ultraviolet radiation (Lopez-Martinez et al. 2008).

The third enzyme modeled is a very-long-chain-acyl-CoA dehydrogenase. These enzymes catalyze the initial step of fatty acid β -oxidation in mitochondria. The β -oxidation of fatty acids is a key process in maintaining energy levels in overwintering insects (Marshall and Sinclair 2012; Sinclair and Marshall 2018); therefore, modifications in or around the ligand binding site may influence the efficiency of this process, whereas surface residue modifications may affect the thermostability (Fields 2001).

The site selection test results we present here point toward selection acting on metabolic associated processes, such as ATP production and the processing of free-radicals. Adaptation to alpine environments requires an organism to be able to deal with stress-induced responses, such as the production of ROS; therefore, the enzymes and their

associated selection patterns represent interesting candidates for further analysis, including functional tests to examine whether any of these amino acid changes affects enzyme function.

Purifying Selection among Wētā Lineages

The overall selection patterns among individual lineages highlight functional constraint. Among the individual internal and terminal branches, between 0 and 307 genes showed significant different ω values, relative to the rest of the phylogeny. Closer investigation of the ω values revealed that although the numbers differ between branches, the trend is toward functional constraint, with $\omega < 1$.

Conclusion

Using transcriptome sequencing we have revealed that metabolic genes tend to be enriched among those under positive selection across the wētā phylogeny. The candidate genes under positive selection include those involved in respiration, energy production, and metabolism. This finding is consistent with the observation that metabolic rate varies among some wētā species (King and Sinclair 2015). Further investigation into the proteins with the most codons under positive selection identified three enzymes with homologous 3D protein structures, allowing the determination of the location of these sites in the overall structure. For the three enzymes, identified as prophenoloxidase, superoxide dismutase, and very-long-chain-acyl-CoA dehydrogenase, the majority of sites are surface residues. These surface modifications are key candidates for further functional tests, as they influence the thermostability and catalytic performance of enzymes. Additionally, prophenoloxidase and very-long-chain-acyl-CoA dehydrogenase have sites near the active site and ligand binding pocket, respectively, identified as under positive selection. These combined results highlight interesting patterns of selection on metabolic loci and protein evolution across the wētā phylogeny.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank R. Cole, M. McDonald, K. Grayson, K. Lomas, D. Seldon, C. Watts, T. Jewell, L. Moran, W. Chin, R. Leschen, I. Millar, A. Walker, P. Ritchie, R. Van Heugten, D. Wharton, M. Lusk, I. Twort, G.A. Twort, O. Twort, and G. Twort for assistance with field work and providing samples and the New Zealand eScience Infrastructure (NeSI) high-performance computing facilities and the staff at the Centre for eResearch at the University of Auckland, for bioinformatic

support. We also thank the iwi from around the country for giving permission for the collection of samples, especially Ngāti Manuhiri and Ngātiwai. We also thank the Department of Conservation for issuing of collecting permits. This work was supported by the Allan Wilson Centre (Grant No. RM13790/24947); The University of Auckland; and Core funding for Crown Research Institutes from the New Zealand Ministry of Business, Innovation and Employment's Science and Innovation Group. We thank Neil Gemmell and three anonymous reviewers for comments that improved the manuscript.

Literature Cited

- Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Mol Biol Evol.* 31(10):2553–2556.
- Abernathy J, Overturf K. 2016. Comparison of ribosomal RNA removal methods for transcriptome sequencing workflows in teleost fish. *Anim Biotechnol.* 27(1):60–65.
- Addo-Bediako A, Chown SL, Gaston KJ. 2002. Metabolic cold adaptation in insects: a large-scale perspective. *Funct Ecol.* 16(3):332–338.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389–3402.
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Barnes AJ, Siva-Jothy MT. 2000. Density-dependent prophylaxis in the mealworm beetle *Tenebrio molitor* L. (Coleoptera: Tenebrionidae): cuticular melanization is an indicator of investment in immunity. *Proc R Soc Lond B Biol Sci.* 267(1439):177–182.
- Berman HM, et al. 2000. The protein data bank. *Nucleic Acids Res.* 28(1):235–242.
- Bulgarella M, Trewick SA, Godfrey AJR, Sinclair BJ, Morgan-Richards M. 2015. Elevational variation in adult body size and growth rate but not in metabolic rate in the tree weta *Hemideina crassidens*. *J Insect Physiol.* 75:30–38.
- Cheviron ZA, Bachman GC, Connaty AD, McClelland GB, Storz JF. 2012. Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proc Natl Acad Sci U S A.* 109(22):8635–8640.
- Cheviron ZA, Whitehead A, Brumfield RT. 2008. Transcriptomic variation and plasticity in rufous-collared sparrows (*Zonotrichia capensis*) along an altitudinal gradient. *Mol Ecol.* 17(20):4556–4569.
- Chown SL, Gaston KJ. 1999. Exploring links between physiology and ecology at macro-scales: the role of respiratory metabolism in insects. *Biol Rev.* 74(1):87–120.
- Clark MS, Worland MR. 2008. How insects survive the cold: molecular mechanisms—a review. *J Comp Physiol B.* 178(8):917–933.
- Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.
- Dalziel AC, Moore SE, Moyes CD. 2005. Mitochondrial enzyme content in the muscles of high-performance fish: evolution and variation among fiber types. *Am J Physiol Regul Integr Comp Physiol.* 288(1):R163–R172.
- Dalziel AC, Rogers SM, Schulte PM. 2009. Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. *Mol Ecol.* 18(24):4997–5017.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 9(8):772.
- Denlinger DL, Lee RE. 2010. *Low temperature biology of insects.* Cambridge: Cambridge University Press.
- Dennis AB, Dunning LT, Sinclair BJ, Buckley TR. 2015. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. *Sci Rep.* 5:13965.
- Dunning LT, Dennis AB, Sinclair BJ, Newcomb RD, Buckley TR. 2014. Divergent transcriptional responses to low temperature among populations of alpine and lowland species of New Zealand stick insects (*Micrarchus*). *Mol Ecol.* 23(11):2712–2726.
- Dunning LT, et al. 2013. Identification of cold-responsive genes in a New Zealand alpine stick insect using RNA-Seq. *Comp Biochem Physiol Part D Genomics Proteomics.* 8:24–31.
- Field LH, Sandlant GR. 2001. The biology of wetas, king crickets and their allies. In: Field LH, editor. *The gallery-related ecology of New Zealand tree wetas *Hemideina femorata* and *Hemideina crassidens** (Orthoptera, Anostomatidae). Wallingford (United Kingdom): CAB International. pp. 243–258
- Fields PA. 2001. Review: protein function at thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol Part A Mol Integr Physiol.* 129(2–3):417–431.
- Fields PA, Somero GN. 1998. Hot spots in cold adaptation: localized increases in conformational flexibility in lactate dehydrogenase A4 orthologs of Antarctic notothenioid fishes. *Proc Natl Acad Sci U S A.* 95(19):11476–11481.
- Foote AD, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet.* 47(3):272–275.
- Fridovich I. 1975. Superoxide dismutases. *Annu Rev Biochem.* 44(1):147–159.
- Fu M, et al. 2011. Autocrine motility factor/phosphoglucose isomerase regulates ER stress and cell death through control of ER calcium release. *Cell Death Differ.* 18(6):1057–1070.
- Gallant JR, et al. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344(6191):1522–1525.
- Ge R-L, et al. 2013. Draft genome sequence of the Tibetan antelope. *Nat Commun.* 4:1858.
- Gene Ontology Consortium. 2015. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43:D1049–D1056.
- Gibbs G. 1998. *New Zealand weta.* Singapore: Kyodo Printing Co Ltd.
- Gibbs G. 2001. The biology of wetas, king crickets and their allies. In: Field LH, editor. *Habitats and biogeography of New Zealand's deinacridine and tusked weta species.* Wallingford (United Kingdom)/New York: CABI Publishers. p. 35–55.
- González-Santoyo I, Córdoba-Aguilar A. 2012. Phenoloxidase: a key component of the insect immune system. *Entomol Exp Appl.* 142(1):1–16.
- Grabherr MG, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652.
- Gretschner RR, et al. 2016. A common theme in extracellular fluids of beetles: extracellular superoxide dismutases crucial for balancing ROS in response to microbial challenge. *Sci Rep.* 6:24082.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52(5):696–704.
- Haas BJ, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8(8):1494–1512.
- Hochachka PW, Somero GN. 2002. *Biochemical adaptation: mechanism and process in physiological evolution.* New York: Oxford University Press.

- Honaas LA, et al. 2016. Selecting superior *de novo* transcriptome assemblies: lessons learned by leveraging the best plant genome. *PLoS One* 11(1):e0146062.
- Hu Y, et al. 2017. Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci USA*. 114(5):1081–1086.
- Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1):27–30.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44(D1):D457–D462.
- Katoh K, Kuma K-I, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33(2):511–518.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kim J-W, Dang CV. 2005. Multifaceted roles of glycolytic enzymes. *Trends Biochem Sci.* 30(3):142–150.
- King K, Sinclair BJ. 2015. Water loss in tree weta (*Hemideina*): adaptation to the montane environment and a test of the melanisation–desiccation resistance hypothesis. *J Exp Biol.* 218(13):1995–2004.
- Kostal V, Renault D, Mehrabianova A, Bastl J. 2007. Insect cold tolerance and repair of chill-injury at fluctuating thermal regimes: role of ion homeostasis. *Comp Biochem Physiol A Mol Integr Physiol.* 147:231–238.
- Lalouette L, Williams CM, Hervant F, Sinclair BJ, Renault D. 2011. Metabolic rate and oxidative stress in insects exposed to low temperature thermal fluctuations. *Comp Biochem Physiol A Mol Integr Physiol.* 158(2):229–234.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
- Li W-H. 1997. *Molecular evolution*. Sunderland (MA): Sinauer Associates.
- Lopez-Martinez G, Elnitsky MA, Benoit JB, Lee RE, Denlinger DL. 2008. High resistance to oxidative damage in the Antarctic midge *Belgica antarctica*, and developmentally linked expression of genes encoding superoxide dismutase, catalase and heat shock proteins. *Insect Biochem Mol Biol.* 38(8):796–804.
- Marden JH. 2013. Nature's inordinate fondness for metabolic enzymes: why metabolic enzyme loci are so frequently targets of selection. *Mol Ecol.* 22(23):5743–5764.
- Marshall KE, Sinclair BJ. 2012. Threshold temperatures mediate the impact of reduced snow cover on overwintering freeze-tolerant caterpillars. *Naturwissenschaften* 99(1):33–41.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17(1):10–12.
- McDonald AG, Boyce S, Tipton KF. 2009. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.* 37(Database):D593–D597.
- Minards NA, Treweek SA, Godfrey AJR, Morgan-Richards M. 2014. Convergent local adaptation in size and growth rate but not metabolic rate in a pair of parapatric Orthoptera species. *Biol J Linn Soc Lond.* 113(1):123–135.
- Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31(12):i44–i52.
- Mishmar D, et al. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A.* 100(1):171–176.
- Morgan-Richards M, Gibbs G. 2001. A phylogenetic analysis of New Zealand giant and tree weta (Orthoptera: Anostostomatidae: *Deinacrida* and *Hemideina*) using morphological and genetic characters. *Invertebr Syst.* 15(1):1–12.
- Nilsson J, Grahm M, Wright AP. 2011. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol.* 12(7):R65.
- Oikawa A, Mori N, Kimura MT. 2006. Comparison of oxygen consumption in drosophilid flies from different climates. *Entomol Science.* 9(4):347–354.
- Overgaard J, Sorensen JG, Petersen SO, Loeschcke V, Holmstrup M. 2005. Changes in membrane lipid composition following rapid cold hardening in *Drosophila melanogaster*. *J Insect Physiol.* 51(11):1173–1182.
- Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.
- Qin W, Neal SJ, Robertson RM, Westwood JT, Walker VK. 2005. Cold hardening and transcriptional change in *Drosophila melanogaster*. *Insect Mol Biol.* 14(6):607–613.
- Qiu Q, et al. 2012. The yak genome and adaptation to life at high altitude. *Nat Genet.* 44(8):946–949.
- Reeson AF, Wilson K, Gunn A, Hails RS, Goulson D. 1998. Baculovirus resistance in the noctuid *Spodoptera exempta* is phenotypically plastic and responds to population density. *Proc R Soc Lond B Biol Sci.* 265(1407):1787–1791.
- Robb T, Forbes MR, Jamieson IG. 2003. Greater cuticular melanism is not associated with greater immunogenic response in adults of the polymorphic mountain stone weta, *Hemideina maori*. *Ecol Entomol.* 28(6):738–746.
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC. 2004. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303(5655):223–226.
- Runck AM, Weber RE, Fago A, Storz JF. 2010. Evolutionary and functional properties of a two-locus β -globin polymorphism in Indian house mice. *Genetics* 184(4):1121–1131.
- Schlessinger A, et al. 2011. Protein disorder—a breakthrough invention of evolution? *Curr Opin Struct Biol.* 21(3):412–418.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.
- Scott GR, et al. 2011. Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol.* 28(1):351–363.
- Seebacher F, et al. 2010. Plasticity of oxidative metabolism in variable climates: molecular mechanisms. *Physiol Biochem Zool.* 83(5):721–732.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Sinclair BJ, Marshall KE. 2018. The many roles of fats in overwintering insects. *J Exp Biol.* 221:Pt Suppl 1.
- Soderhall K, Cerenius L. 1998. Role of the prophenoloxidase-activating system in invertebrate immunity. *Curr Opin Immunol.* 10(1):23–28.
- Storey JD. 2002. A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol.* 64(3):479–498.
- Storey KB, Storey JM. 2012. Insect cold hardiness: metabolic, gene, and protein adaptation. *Can J Zool.* 90:456–475.
- Storz JF, Scott GR, Cheviron ZA. 2010. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J Exp Biol.* 213(24):4125–4136.
- Storz JF, Zera AJ. 2011. Molecular methods for evolutionary genetics. In: Orgogozo V, Rockman MV, editors. *Experimental approaches to evaluate the contributions of candidate protein-coding mutations to phenotypic evolution*. Totowa (NJ): Humana Press. p. 377–396.

- Trewick SA, Morgan-Richards M. 2004. Phylogenetics of New Zealand's tree, giant and tusked weta (Orthoptera: Anostostomatidae): evidence from mitochondrial DNA. *J Orthoptera Res.* 13(2):185–196.
- Trewick SA, Morgan-Richards M. 2005. After the deluge: mitochondrial DNA indicates Miocene radiation and Pliocene adaptation of tree and giant weta (Orthoptera: Anostostomatidae). *J Biogeogr.* 32(2):295–309.
- Twort VG, et al. 2017. Positive selection and comparative molecular evolution of reproductive proteins from New Zealand tree weta (Orthoptera, Hemideina). *PLoS One* 12(11):e0188147.
- Voituron Y, Mouquet N, de Mazancourt C, Clobert J. 2002. To freeze or not to freeze? An evolutionary perspective on the cold-hardiness strategies of overwintering ectotherms. *Am Nat.* 160:255–270.
- Wang Y, Geer LY, Chappay C, Kans JA, Bryant SH. 2000. Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci.* 25(6):300–302.
- Waterhouse A, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46(W1):W296–W303.
- Waterhouse RM, et al. 2007. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 316(5832):1738–1743.
- Wheat CW, et al. 2011. Functional genomics of life history variation in a butterfly metapopulation. *Mol Ecol.* 20(9):1813–1828.
- Wilson K, Cotter SC, Reeson AF, Pell JK. 2001. Melanism and disease resistance in insects. *Ecol Lett.* 4(6):637–649.
- Wilson K, et al. 2002. Coping with crowds: density-dependent disease resistance in desert locusts. *Proc Natl Acad Sci U S A.* 99(8):5471–5475.
- Xie C, et al. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39(Suppl 2):W316–W322.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol.* 28(3):1217–1228.
- Yang Z, Nielsen R, Goldman N, Pedersen A-M. 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics.* 155:431–449.
- Yokoyama S, Tada T, Zhang H, Britt L. 2008. Elucidation of phenotypic adaptations: molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci U S A.* 105(36):13480–13485.
- Zhang F, Broughton RE. 2015. Heterogeneous natural selection on oxidative phosphorylation genes among fishes with extreme high and low aerobic performance. *BMC Evol Biol.* 15:1–15.
- Zhang G, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346(6215):1311–1320.
- Zhang Z, et al. 2016. Convergent evolution of rumen microbiomes in high-altitude mammals. *Curr Biol.* 26(14):1873–1879.
- Zwickl DJ. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Austin (TX): University of Texas at Austin.

Associate editor: David Bryant