2024-08-01

Signatures Of Selection, Gene Flow, And Genetic Drift On The Genomes Of A Recent Avian Radiation

FLOR BRIGITTE HERNANDEZ CAMACHO
University of Texas at El Paso

SIGNATURES OF SELECTION, GENE FLOW, AND GENETIC DRIFT ON THE GENOMES OF A RECENT AVIAN RADIATION

FLOR BRIGITTE HERNANDEZ CAMACHO

Doctoral Program in Ecology and Evolutionary Biology

APPROVED:

Philip Lavretsky, Ph.D., Chair

Markus J. Peterson, Ph.D.

Eli Greenbaum, Ph.D.

Ming-Ying Leung, Ph.D.

Caitlin Wells, Ph.D.

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

By

Flor B. Hernandez Camacho

2024

SIGNATURES OF SELECTION, GENE FLOW, AND GENETIC DRIFT ON THE GENOMES OF A RECENT AVIAN RADIATION

by

FLOR BRIGITTE HERNANDEZ CAMACHO, B.S

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Biological Science

THE UNIVERSITY OF TEXAS AT EL PASO

August 2024

Acknowledgments

I would like to express my deep gratitude to my dissertation advisor, Philip Lavretsky, for giving me the opportunity to be part of his research team, his constant confidence in me, and his direct and honest support and guidance. I am also grateful for my committee members for their guidance and for helping me grow as a researcher.

I would like to thank the Lavretsky lab's team members for their constant support, both personally and academically: Josh I. Brown, Sara Gonzalez, Lauren Mcfarland, Vergie Musni, Kristi Fukunaga, Marissa Kaminski, and Niko Enriquez. Likewise, I also want to thank faculty, staff, and graduate students in the Ecology and Evolutionary Biology Program. Special thanks to Jonathon Mohl, Mike Harvey, and Oscar Johnson for their guidance in improving my dissertation, and to Elizabeth Walsh, Vanessa Lougheed, and Michael Moody for their great work during my time in the department. Additionally, thank you to the BEES (Biology, Environmental and Engineering Student Group) and the 500 Women Scientist- El Paso Pod.

Thank you also to my mentors, friends, and colleagues for encourage me to pursue a Ph.D. and their support in this journey, especially Thomas Valqui, Luis Alza, Kevin McCracken, Chris Witt, Stu Mackenzie, A. Quinonez, M. Combe, S. Figueroa, K. Verde, C. Landauro, M. Cuyos, C. Apolinario, V. Cueva, C. Martell, M. Ramos, C. Huapaya, and S. Cespedes.

I am very grateful to my Peruvian friends at El Paso, for their love and support, making feel at home. Special thanks to Henry Moncada, Johana Zumaran, Leo Zumaran, Jeanette Orbegozo, Pedro Palermo, Erik Fernandez, and Noraya Ccoyure. I also thank my family in Peru — my parents Flora Camacho, Elvis Hernandez, and my brother Elvis Hernandez — for their love and support. Finally, I am profoundly thankful to my husband, Alexis Diaz, for his immense love, support, and guidance throughout these years.

Abstract

Speciation is not a discrete event but rather a gradual process that occurs along a continuum. Initially, populations remain genetically and reproductively connected, but they become reproductively isolated as the process of speciation progresses. Advances in DNA sequencing technology has helped our understanding of the processes that contribute to changes in the genetic composition of populations over the progression of speciation, also known as drivers of evolution (i.e. genetic drift, mutation, selection, and gene flow). Among them, hybridization, or the interbreeding between distinct lineages, is increasingly recognized as a natural process that can have significant consequences to the speciation process. In particular, the frequency and directionality of introgressive hybridization (i.e., gene flow) dictates the extent of its impact(s) on the evolutionary trajectories. Importantly, the proximate result from gene flow is also tied to the directionality and strength of natural selection, as well as intensity of genetic drift. Here, I test hypotheses surrounding early species divergence, including the influence of these different evolutionary mechanisms on the process. Using a whole-genome sequencing approach, I aim to understand the genetic and evolutionary consequences of natural and human-mediated gene flow in a recently radiated avian system, the Mallard Complex. Importantly, I use inferences made from my study system to also inform conservation practices.

In Chapter 1, I focused on the genomic consequences of hybrid speciation by comparing the full genomes of a putative young homoploid hybrid species, the Hawaiian duck (*Anas wyvilliana*). I conducted full genome comparison of the Hawaiian duck to its parental species, the island endemic Laysan duck (*Anas layanensis*) and a mainland generalist mallard (*Anas platyrhynchos*). I found that the Hawaiian duck's genome is indeed a mosaic of genetic ancestry from both parental taxa, with a predominant contribution from the Laysan duck. Although the

extent of reproductive isolation from either parental species is still unknown, I found potential genes associated with reproductive barriers, particularly on the Z-sex chromosome, related to fertilization, male courtship, and embryo development. Overall, my results are consistent with the hypothesis that the Hawaiian duck evolved via hybrid speciation and shed light on genes potentially advantageous for the emergence and persistence of this nascent hybrid species in the Hawaiian Islands.

In Chapter two, I continue to focus on the Hawaiian duck that is currently threatened by introgressive hybridization from domestic mallards, highlighting the pressing conservation concern of potential genomic extinction and loss of adaptiveness for native species because of the extensive introgression of non-native genes. To alleviate or reverse trends for such scenarios requires the direct integration of genomic data within a model framework for effective management. Towards this end, I developed the simRestore R program as a decision-making tool that integrates ecological and genomic information to simulate ancestry outcomes from optimized conservation strategies. The program optimizes supplementation and removal strategies across generations until a set native genetic threshold is reached within the studied population. Importantly, in addition to helping with initial decision-making, simulations can be updated with the outcomes of ongoing efforts, allowing for the adaptive management of populations. After demonstrating functionality, I apply and optimize among actionable management strategies for the endangered Hawaiian duck for which the current primary threat is genetic extinction through ongoing anthropogenic hybridization with feral mallards. Simulations demonstrate that supplemental and removal efforts can be strategically tailored to move the genetic ancestry of Hawaii's hybrid populations towards Hawaiian duck without completely starting over. Further, I discuss ecological parameter sensitivity, including which factors are

most important to ensure genetic outcomes (i.e. number of offspring). Finally, to facilitate use, the program is also available online as a Shiny Web application.

My third Chapter focuses on understanding the diverse signals present in the species' genomes that reflect the interplay of genetic drift, selection, and gene flow during the speciation process. Although unraveling the adaptive histories of species often includes determining how directional selection uniquely acts across respective genomes, it is equally important to establish what, if any, parts remain under purifying selection for ancestral state(s). Yet, this is only possible in sufficiently divergent genomes, allowing for the identification of both genomic islands and evolutionary valleys. To this end, my research involves comparing the full genomes of 11 of 14 species within the Mallard Complex, which represent successful adaptive radiation and the speciation continuum, to understand how directional and purifying selection have shaped derived and ancestral states, respectively. First, phylogenetic and demographic analyses support the "out of Africa" hypothesis of the Mallard Complex, dating back between 1 and 2 million years, with recent divergences occurring in the last 500,000 years among North American species. Second, while genomic islands of differentiation were identified in 44 pairwise species comparisons representing early and moderate stages of divergence, genomic valleys of ancestral retention were observed in all comparisons representing all stages of divergence. Finally, surveying gene ontology (GO) terms revealed a striking pattern in which valleys of similarity are linked to fundamental organismal functions and survival, and which outnumber islands of differentiation associated with adaptation to environmental challenges. Overall, my findings underscore the importance of analyzing species representing the speciation continuum to identify regions under divergent and purifying selection and discern potential stochastic mirages because of genetic drift.

Finally, Chapter four examines the consequences of domestic introgression into the genomes of their wild counterparts using the Hawaiian duck, New Zealand grey duck (*Anas superciliosa*), North American mallard and their respectively domestic breeds as study system. Using whole-genome re-sequencing, I analyzed domestic and wild congeners, along with resulting hybrids representing different levels of admixture across these regions. Assessing genetic diversity, runs of homozygosity, and demographic parameters for each genome, I consistently found that wild genomes exhibited higher genetic diversity, lower runs of homozygosity, and reduced inbreeding coefficients compared to their domestic counterparts. Across ecoregions, the strongest statistical correlation between ancestry and summary statistics was observed in North America, where domestic game-farm mallards are regularly introduced and interact with wild populations. Conversely, in regions like New Zealand and Hawaii, where stocking efforts have declined or populations are small, I identified evidence of adaptive selection and genetic drift as dominant forces shaping genomes. Furthermore, I illustrated how demographic history inferences using genomes derived from domestication or hybrid origins can be severely biased, leading to distorted estimates of effective population size (N_E) and divergence times. Ultimately, I concluded that relying on highly inbred individuals to supplement populations of wild conspecifics not only exacerbates hybridization but might also intensify inbreeding, leading to important adaptive and conservation implications.

Overall, my dissertation investigates the intricate dynamics of speciation within the Mallard Complex, emphasizing the continuous nature of this evolutionary process. By recognizing speciation as a gradual continuum rather than discreet event(s), I have uncovered the

multifaceted roles of natural selection, genetic drift, and gene flow in driving genetic divergence and ultimately species formation. Through meticulous genomic analyses and simulations, I have elucidated the genetic and evolutionary consequences of hybridization. Importantly, my work resulted in the development of a novel decision-making tool, simRestore, which integrates ecological and genomic data to inform conservation strategies for threatened species. Ultimately, this research underscores the importance of studying the dynamics of hybridization at both evolutionary and contemporary scales to advance our knowledge of speciation mechanisms and their implications for the conservation of Biodiversity.

Table of Contents

List of Tables

List of Figures

Chapter 1: Whole genome sequence analysis supports the genomic mosaicism of the endangered Hawaiian duck

ABSTRACT

Once considered rare events, hybridization and gene flow have been determined as pervasive evolutionary processes across much of biodiversity, leading to research foci attempting to understand proximate evolutionary outcomes from such events. Here, I compare the full genomes of a putative young homoploid hybrid species, the Hawaiian duck, to its parental species, the Hawaiian Island endemic Laysan duck and a mainland generalist mallard. I find that the Hawaiian duck's genome is indeed a mosaic of genetic ancestry of the two parental taxa, with a higher contribution coming from Laysan duck than mallard, but also identify genomic regions harboring Hawaiian duck-derived alleles. Although the extent of reproductive isolation from either parental species is still unknown, I find putative candidate genes related to reproductive barriers such as fertilization, male courtship behavior, and embryo development on the Z-sex chromosome. Importantly, I highlight genomic regions that harbor putative genes for species-specific adaptation including immune system, appetite, and fear response. Overall, my results support a hybrid origin for the Hawaiian duck and demarcate potential genes that may have been evolutionarily beneficial for the rise and persistence of this young hybrid species on the main Hawaiian Islands. Finally, my study resulted in the first reference genomes for the endangered Hawaiian and Laysan ducks, which will serve as important tools for their future conservation efforts.

INTRODUCTION

Hybridization is defined as the interbreeding of individuals from genetically distinct populations irrespective of taxonomic status (R. Abbott et al., 2013a; Grant & Grant, 1992a; J. Rhymer, 2006). Importantly, hybridization resulting in gene flow where genetic material moves between distinct organisms and thereby modifying gene pool(s) (J. Rhymer, 2006) is increasingly being recognized as an important and pervasive evolutionary process (Mallet, 2007a; Rheindt & Edwards, 2011). In particular, the potential of interspecific interactions resulting in evolutionary adaptive scenarios or even in the speciation of novel taxa not only depends on the strength of reproductive barriers between the interbreeding species (R. Abbott et al., 2013a; R. J. Abbott et al., 2010; Barton, 2013a; Mallet, 2007a), but also on the fitness level of the hybrids (Nolte & Tautz, 2010a). In fact, if a hybrid lineage is better fit than their parental lineages in a new ecological niche, and subsequent reproductive isolation between hybrids and parents is achieved, then a separate species may arise (Mallet, 2005a; Rieseberg & Willis, 2007). Examples of natural hybrid species that evolve via hybrid speciation are rare, but have been identified across plants (Rieseberg et al., 1999), insects (Mavárez et al., 2006), flies (Schwarz et al., 2005), mammals (Chafin et al., 2020), and birds (Lamichhaney et al., 2018). In addition to hybrid speciation, signatures of genomic admixture can arise from the selection of one or few loci that influence a unique adaptive trait via adaptive introgression (R. Abbott et al., 2013a; R. J. Abbott et al., 2010; Jiggins et al., 2008; Schumer et al., 2014). Distinguishing between these scenarios is essential when attempting to understand their prevalence and importance on the evolution of species' genomes versus the potential for novel species formation (Jiggins et al., 2008).

Order Anseriformes presents the most extensive cases of hybridization, with an incidence around 40 – 60% (Aliabadian & Nijman, 2007; Grant & Grant, 1992a). In particular, the Mallard complex (Genus *Anas*), which is a group of 13-14 closely related taxa of waterfowl distributed around the world (depending on taxonomic authority: Johnsgard, 1960; Lavretsky et al., 2014), show high levels of hybridization because of their recent ancestry and/or lack of pre-zygotic barriers (Lavretsky et al., 2014). Among these species, the Hawaiian duck (*Anas wyvilliana*) constitutes a hypothetical example of hybrid speciation because of an ancient hybridization event between the island endemic Laysan duck (*Anas layanensis*) and the broadly distributed mallard duck (*Anas platyrhynchos*) (Lavretsky et al., 2015). The evolutionary history of the Hawaiian duck was difficult to reconcile because of incongruencies in sister relationships reconstructed with different markers or traits. Whereas morphology (Livezey 1993) and nuclear DNA identifify the Hawaiian duck as sister to Laysan ducks, mitochondrial DNA puts it closer to mallards (Lavretsky et al., 2014, 2015). In fact, through additional individual and genetic sampling, Lavretsky et al. (2015) was able to confirm that Hawaiian ducks are best described as a young hybrid species, which resulted from a relatively recent (~1,000-5,000 years before present) but ancient hybridization event between mallards and Laysan ducks. Thus, the genome of the Hawaiian duck offers a unique opportunity to gain insights into the intricate process of genome admixture within a nascent hybrid lineage, while also testing whether such genomes arise because of hybrid speciation or adaptive introgression.

The endangered Hawaiian duck is endemic to the main Hawaiian Islands, and it is phenotypically and ecologically distinct from all other mallard-like ducks. However, this species can be considered phenotypically intermediate, in terms of morphology and plumage, with respect to its putative parental species (i.e. Laysan duck and mallard) (Engilis Jr. et al., 2020a;

Lavretsky, 2021; Lavretsky et al., 2015; Uyehara et al., 2007). Previous morphological studies have suggested that Hawaiian ducks show intermediate intra-appendicular skeletal and sternal dimensions with respect to its putative parental species (Livezey, 1993). Additionally, plumage characteristics of Hawaiian ducks have also been found to be intermediate, with individuals displaying everything from Laysan duck to mallard-like traits (Engilis Jr. et al., 2020a). These morphological intricacies are similar to other studies of putative hybrid species that often exhibit intermediate phenotypes to their parental species (R. J. Abbott et al., 2010). For example, the *Heliconius heurippa* butterfly is considered a hybrid species that exhibits an intermediate wing color pattern that promotes an assortative mating preference among hybrids (Mavárez et al., 2006; Salazar et al., 2005). Although such morphological traits are easily described, the genomic basis of these remains poorly understood. Towards this end, I compare genomes of the Hawaiian duck and its two putative parental taxa to establish whether behavioral and morphological intermediacy of the Hawaiian duck may be a result of a genomic mosaicism. This condition involves a significant contribution of genetic material from both parental lineages, and it is considered a key signature of hybrid speciation (Schumer et al., 2014, Elgvin et al., 2017). I present the first reference Hawaiian and Laysan duck genomes, and combined these with the recent mallard reference genome (Lavretsky et al., 2023a), and test whether the Hawaiian duck's genome is indeed a mosaic of Laysan duck and mallard genetic diversity as predicted under a hybrid speciation scenario. Alternatively, the Hawaiian duck may simply be a case of adaptive introgression in which the genome is largely Laysan duck-derived with the exception of a few mallard-derived genes. Additionally, I determine whether Hawaiian ducks have been sufficiently isolated from its parental taxa to have derived novel ancestry since divergence. Finally, I delve into potential functionality within genomic regions that remain mallard and Laysan duck-

derived, as well as those that are unique to Hawaiian ducks within the Hawaiian duck's genome. I predict Hawaiian ducks to harbor Laysan duck ancestry associated with traits adapted to Hawaiian Island ecology, while mallard-derived genetic variation to be linked to predator evasion and/or other traits that may explain how this duck survived Polynesian colonization while so many other Island species did not (Athens et al., 2002; Boyer, 2008).

METHODS

Sampling, DNA extraction, whole-genome resequencing, and variant calling.

Tissue samples were collected from one Laysan duck (May 2016, Hawaiian Islands), Hawaiian duck (November 2016, Hawaiian Islands), and mallard (September 2016, New Mexico) (Supplementary Table S1.1). Wing or breast tissue was sent to Cantata Bio, LLC (Scotts Valley, CA) for DNA extraction, library preparation, and sequencing.

Genome sequencing, assembly, and annotation

All steps were done at Cantata Bio laboratories. DNA was first fragmented using the Bioruptor Pico for Illumina library creation. Two short-insert libraries were produced, one with an approximate insert size of 400 bp and the other with 500 bp, following Cantata Bio's protocol and the Illumina TruSeq DNA PCR-free method. These libraries underwent sequencing on an Illumina HiSeq machine using paired-end (PE) 150-bp chemistry. Following, raw Illumina reads were first trimmed to remove sequencing adapters and poor-quality bases using Trimmonatic (Bolger et al. 2014). The short-insert sequences underwent analysis at various k-mer sizes (19, 31, 49, 75, 109), with a negative binomial model applied to optimize the assembly's coverage and to balance repetitive and heterozygous sequences. De novo assembly was performed using Meraculous v2 (Chapman et al. 2011), setting the k-mer size at 55 and requiring a minimum k-mer frequency of 15, while employing the diploid nonredundant haplotig mode.

Following de novo assembly with Meraculus, three Chicago libraries were prepared (Cantata Bio) following methods outlined in Putnam et al. (2016). Briefly, around 500 ng of HMW DNA underwent in vitro chromatin reconstitution and was then fixed using formaldehyde. The chromatin was treated with DpnII, with biotinylated nucleotides filling in the 5' overhangs, and the remaining blunt ends were ligated. Post-ligation, crosslinks were undone to separate proteins from the DNA. The purified DNA underwent a process to eliminate non-ligated biotin, followed by fragmentation to an average size of approximately 350 bp. Sequencing libraries were created from this fragmented DNA using NEBNext Ultra enzymes and adapters compatible with Illumina. Biotin-labeled fragments were extracted using streptavidin beads prior to PCR amplification. Libraries were each sequenced on a single Illumina HiSeq X Lane using 150 bp PE chemistry.

Subsequently, to scaffold and improve assembly, shotgun and Chicago library data was combined with the initial assembly from Meraculous as the input for the HiRise computational pipeline (Putnam et al., 2016). First, the shotgun and Chicago library sequences were aligned to the draft contig assembly using Burrows Wheeler Aligner v07.15 (bwa, Li & Durbin, 2009). Second, the Chicago read pairs were mapped within draft scaffolds to produce a likelihood model for genomic distance between read pairs, and this model was used to identify and break putative misjoins, score prospective joins, and make joins above the threshold. Finally, after aligning and scaffolding the draft assembly using the Chicago data, the short-insert sequences were used to close remaining gaps between contigs where possible. For the mallard duck genome, Omni-C libraries were meticulously prepared to enhance the genome assembly process (see Lavretsky et al., 2023). For Laysan and Hawaiian duck genomes, I assessed contiguity using

Quast (Gurevich et al., 2013), and used BUSCO v3.1.0 (Waterhouse et al., 2018) with the associated eukaryota odb9 dataset to attain BUSCO scores.

Finally, genomic assemblies were passed through Omics-box (https://www.biobam.com) to identify protein coding regions using Augustus (Hoff & Stanke, 2019). OrthoVenn2 (Xu et al., 2019) was employed to cluster homologous proteins and generate Venn diagrams across the three duck genomes.

Variant calling

Variant calling was performed across raw FASTQ files representing the three ducks (Laysan duck, Hawaiian duck and mallard) using a custom in-house Python script (Python scripts available at https://github.com/jonmohl/PopGen; Lavretsky et al., 2020). Trimmomatic (Bolger et al., 2014) was used to trim or discard poor quality sequences. Quality sequence reads were then aligned to a high-quality de novo reference genome of a wild North American mallard (Lavretsky et al., 2023; Available at: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030704485.1/) using the Burrows Wheeler Aligner v07.15 (bwa; Li & Durbin, 2009). Samples were then sorted and indexed in Samtools v1.6 (Li et al., 2009) and combined and genotyped using bcftools v1.6 (as part of the SAMtools package) "mpileup" and "call" functions with the following parameters "-c –A -Q 30 -q 30," which set a base pair and overall sequence PHRED score of ≥30 to ensure that only high-quality sequences are retained.

Summary statistics

I assessed genome-wide heterogeneity by estimating pair-wise species relative divergence (F_{ST}) and absolute divergence (D_{XY}), as well as per species nucleotide diversity (π) based on non-overlapping 100kb window sizes and using the Python script, popgenWindows.py (Martin et al., 2015; available in https://github.com/simonhmartin/genomics_general). Additionally, runs of homozygosity (ROH) were also calculated for each genome using VCFtools v0.1.17 (Danecek et al., 2011) and using all possible base-pairs.

Ancestry painting

I first visualized the Hawaiian duck's ancestry blocks through ancestry painting following Barth et al., (2020) and Runemark et al., (2018; https://github.com/millanek/tutorials/tree/master/analysis_of_introgression_with_snp_data). I first identified fixed genomic differences between the Laysan duck and mallard with no missing data using the script get_fixed_site_gts.rb. Following, base-pairs were thinned to be a minimum of 1kb apart. I plotted the ancestry-informative sites in the Hawaiian duck using the script plot_fixed_site_gts.rb. If Hawaiian ducks represent a hybrid species, then I expect them to have a mosaicism of Laysan duck and mallard ancestry as previously found with a few loci (Lavretsky et al., 2015). Note that all samples were males (genotype = ZZ), and thus the Z-sex chromosome was treated as diploid like autosomes. Although the mitogenome is haploid, this was also treated as diploid, and following Elgvin et al., (2017).

Patterns of gene flow and ancestry blocks

I used a four-taxon ABBA-BABA (D' statistics) calculations to test for the presence of introgression (Durand et al., 2011; Green et al., 2010). The Tufted duck (*Aythya fulvigula*) genome (retrieved from NCBI, https://www.ncbi.nlm.nih.gov/genome/?term=Aythya+fuligula; Kraus et al. Unpublished) was aligned to my three genomes and served as the outgroup in analyses. Whole genome ABBA-BABA estimates and jack-knifed standard errors were calculated for two topologies that included: {(Laysan, Hawaiian), mallard [Tufted]} and {(mallard, Hawaiian), Laysan [Tufted]}, using R code detailed in Martin et al., (2015; https://github.com/simonhmartin/tutorials/tree/master/ABBA_BABA_whole_genome). In addition to calculating the D' statistics, I quantified the f_d estimator using the ABBABABAwindow.py script (Martin et al., 2015), and with the Tufted duck serving as an outgroup again. Both analyses were based on non-overlapping 100kb windows, with a minimum requirement of 10 SNPs per window.

Finally, I assessed the sizes of ancestry blocks inherited from distinct parental species within the Hawaiian duck following Meier et al. (2017; also see Runemark et al., 2018). To do so, I estimated the frequency of ABBA and BABA sites within non-overlapping 3kb windows. Windows in which the ratio of ABBA/(ABBA+BABA) exceeded 0.7 were identified as potential Laysan duck ancestry blocks, whereas those with a ratio below 0.3 were considered potential mallard ancestry windows within the Hawaiian duck genome. The length of each ancestry block was determined by the number of consecutive windows exhibiting the same ancestry pattern.

Phylogenetic analysis

I took a phylogenetic approach to understand the relationships between the Hawaiian duck and its parental species, with the Tufted duck serving as the outgroup. First, the full VCF file was thinned to include only SNPs which were 100 bp apart using VCFtools v0.1.17 (Danecek et al., 2011; -thin 100). The SNP dataset was further filtered using bcftools (part of SAMtools v1.6; bcftools view -i 'COUNT(GT="RR") > 0 & COUNT(GT="AA")>0') to leave only SNPs that were present at least once in my dataset as homozygous for the reference allele, and homozygous for the alternative allele, as required by RAxML v8.026 (Stamatakis, 2014).The reduced VCF file was then partitioned into two dataset containing autosomal or Z-sex chromosome linked SNPs. For each of these datasets, VCF files were then converted to a fasta file using vcf2phylip.py (Ortiz, 2019) to generate alignments. To investigate possible variations in coalescent histories between different genomic segments, I divided the fasta files into non-overlapping windows of 100 bp. Within each window, I estimated maximum likelihood trees using RAxML v8.026 (Stamatakis, 2014), employing the GTR-GAMMA model. For each possible subtree insertion, RAxML evaluates the average node placement distance of the alternative topology measuring how well a subtree fits into larger phylogenetic context. Window-based subtrees were analyzed for whether the Hawaiian duck was sister to Laysan duck, mallard, or formed a unique clade.

Demographic history

Demographic histories for Hawaiian ducks, Laysan ducks, and mallards were reconstructed using the Pairwise sequential Markovian Coalescent (PSMC) models (Li & Durbin, 2011; https://github.com/lh3/psmc). I first filtered each genome for a minimum read

sequencing depth of 10, and then applied PSMC parameters optimized for avian genomes as outlined in Nadachowska-Brzyska et al., (2015). These parameters included a maximum number of iterations (N = 30), a maximum 2N0 coalescent time (t = 5), an initial theta/rho ratio (r = 5), and a specific pattern of parameters (p = "4+30*2+4+6+10"). Each analysis was executed with 100 bootstrap replicates to ensure robustness. Finally, I converted the PSMC parameter into biologically informative values using a generation time (G) calculated as $G = \alpha + (s / (1 - s))$, where α is the age of maturity and s is the expected adult survival rate (Sæther et al., 2005). For mallard-like ducks, the age of maturity is typically 1 year (i.e. α = 1; Alerstam & Högstedt, 1982). However, I estimated the average survival rate individually. For wild mallards, this survival rate is ~0.57 (i.e. range: 0.46–0.68; Arnold & Clark, 1996; Drilling et al., 2020; Smith & Reynolds, 1992); resulting in an estimated generation time of 2.32 years. For Layan duck, the survival rate is 0.906 (M. H. Reynolds & Citta, 2007), leading to an estimated generation time of 10.64 years. For Hawaiian ducks, the survival rate is 0.74 (Malachowski et al., 2022), resulting in an estimated generation time of 3.97 years. The nuclear mutation rate was set to 1×10^{-9} (Lavretsky et al., 2020).

Identification of high-divergence regions and gene analysis

Identifying candidate loci within putative outlier regions is often challenging when working with a hybrid system as such lineages inherit evolutionary histories from two sources that can distort signals (Elgvin et al., 2017; Runemark et al., 2018). Rather, I used disparities in F_{ST} values between lineages to identify genomic regions where the Hawaiian duck display elevated divergence from either one or both of its parent species. I selected the top 1% 100-kb windows where the Hawaiian duck exhibited the largest difference in F_{ST} values between one parent and

the other per comparison between the Hawaiian duck and either parent, and following these equations (Eq.1 and 2: Figure 1.4A and Supplementary Figure S1.4):

(1) Laysan versus Hawaiian duck divergence (LH) = LH F_{ST} – MH F_{ST}

(2) mallard versus Hawaiian duck divergence (MH) = MH F_{ST} – LH F_{ST}

Next, Hawaiian duck-specific outliers were based on the top 1% of 100kb-windows exhibiting the largest differences in F_{ST} between each hybrid/parent comparison, keeping only windows overlapping between both hybrid-parent comparison as PH window, and following the equation (Eq.3, Figure 1.4D and Supplementary Figure S1.5):

(3) Parents versus Hawaiian duck divergence (PH) = MH F_{ST} – LM F_{ST} ∪ LH F_{ST} – LM F_{ST}

Finally, I investigated for enrichment of Gene Ontology (GO) terms within the high-divergence regions (LH, MH, and PH) by first isolating the genes and associated GO terms within those regions. Next, I used an in-house Python script to determine the GO terms that are significantly over-represented after applying a Bonferroni correction, with an adjusted *p* value cutoff of 0.05. Additionally, metabolic pathways were identified via the KEGG (Kyoto Encyclopedia of Genes and Genomes) database (Kanehisa et al., 2002) using GhostKOALA (Kanehisa et al., 2016; Available at the website: https://www.kegg.jp/ghostkoala/). Then, KEGG enrichment analyses was performed using Enrichr tool (Chen et al., 2013, Available at: https://maayanlab.cloud/Enrichr/) with a Benjamini-Hochberg correction at an adjusted *p* values cutoff of 0.05.

RESULTS

Genome sequencing and assembly results

I constructed scaffold-level reference genomes for the Laysan duck and Hawaiian duck, employing a multi-level approach that combined two sequencing and assembly technologies (shotgun and Chicago; Supplementary Table S1.2). Generating approximately 510 million paired reads for both genomes, I achieved 36-fold and 71-fold coverage genomes for the Laysan and Hawaiian duck assemblies, respectively. Both assembled genomes were ~1.04 Gb with a 41.01% GC content. Notably, both genomes exhibited scaffold N50 of 0.065 Mb, and the longest scaffold length reaching 1,060 Mb (Supplementary Table S1.2). Following the scaffolding of library reads using the Hi-Rise pipeline, the final Laysan duck assembly comprised 9,721 scaffolds covering 1 Gb at the scaffold level, with an N50 sequence size of 0.761 Mb and a BUSCO completeness score of 80.5%. In comparison, the final Hawaiian duck assembly consisted of 3,859 scaffolds encompassing 1 Gb at the scaffold level, with an N50 sequence size of 8.035 Mb, and a BUSCO completeness score of 81.5%. This is compared to the recently published 444.6-fold coverage mallard genome that was ~1.04 Gb with a slightly lower GC content of 40.9%, but with a BUSCO completeness score of 95.4% (Lavretsky et al., 2023b).

The total number of predicted proteins varied by genome, with the highest count observed in the Laysan duck (24,952), followed by the Hawaiian duck (23,927), and the wild mallard having the lowest number (23,584). Utilizing OrthoVenn2, I identified an overlap of 21,680 protein clusters across the three ducks. Notably, the most significant overlap in protein clusters was observed between the Laysan duck and the Hawaiian duck (868), followed by the overlap between the Laysan duck and the mallard (599), and the lowest overlap was observed between

the mallard and the Hawaiian duck (442) (Supplementary Figure S1.1). The number of singletons (i.e., proteins that did not form a cluster and were not included in Supplementary Figure S1.1) also varied across species, being 1,446, 655, and 578 for the Laysan duck, Hawaiian duck, and mallard, respectively. Finally, note the general lack of unique protein clusters within the Hawaiian duck and mallard genomes, and with the Laysan duck possessing only one unique protein cluster (Supplementary Figure S1.1).

Finally, all downstream analyses were based on the alignment of the three and Tufted duck genomes to the wild mallard chromosome-level reference genome, permitting us to partition datasets into autosomal, Z-sex chromosome, and the mitogenome. In the end, I attained high alignment similarity across genomes (~99%; Supplementary Table S1.3) for a total of 879,691,675 (of ~1.04 Gbp) quality base-pairs, including 63,805,135 single nucleotide polymorphisms (SNPs).

Summary statistics and genome painting

Average relative genomic differentiation (F_{ST}) values were markedly higher between the parent species (average F_{ST} Laysan duck versus mallard = 0.38) as compared to Hawaiian ducks and either the Laysan duck (average F_{ST} = 0.27) or the mallard (average F_{ST} = 0.11) (Table 1.1). Although Hawaiian ducks and mallards had similarly high nucleotide diversity (avg. $\pi \sim 0.0078$), mallards had the shortest runs of homozygosity, followed by Hawaiian and Laysan ducks (Table 1.1). Moreover, I recovered disparities in heterozygosity levels across genomic regions of the Hawaiian duck, taking into account both putative parent species as fully homozygous. Specifically, the Z-sex chromosome exhibited higher heterozygosity (0.43) as compared to

autosomes (0.35) (Fig 1A; Supplementary Figure S1.2, and Supplementary Table S1.4). In fact, painting Laysan duck and mallard ancestry revealed a genomic mosaicism for the Hawaiian duck (Figure 1.1A). Interestingly, I find that the Hawaiian duck had a greater number of homozygous sites shared with the mallard across autosomes, whereas they shared a greater number of homozygous sites with the Laysan duck across the Z-sex chromosome (Supplementary Table S1.2). Finally, I found that Hawaiian duck harbors the mallard mitogenome (Figure 1.1A), supporting earlier hypotheses of a mitochondrial capture event (Lavretsky et al., 2015).

Table 1.1. Mean values of genome statistics for 100 kb non-overlapping sliding windows and run of homozygosity (ROH).

Parameters	Species	Autosomes	Z chromosome
F_{ST}	LM	0.389 ± 0.115	0.387 ± 0.135
	LH	0.272 ± 0.254	0.137 ± 0.222
	MH	0.116 ± 0.162	0.123 ± 0.168
D_{XY}	LM	0.0096 ± 0.0056	0.0047 ± 0.0035
	LH	0.0064 ± 0.0047	0.0026 ± 0.0025
	MH	0.0091 ± 0.0054	0.0045 ± 0.0034
π	Laysan	0.0009 ± 0.0014	0.0006 ± 0.0005
	Hawaiian	0.007 ± 0.0056	0.0035 ± 0.0034
	mallard	0.0086 ± 0.0049	0.0042 ± 0.0033
ROH	Laysan	20,869 ± 39,676	18,954 ± 26,083
	Hawaiian	14,132 ± 23,750	11,619 ± 15,366
	mallard	9,621 ± 15,133	4,283 ± 11,160

Estimating gene flow and ancestry blocks

My four-taxon ABBA-BABA (D' statistics) test revealed substantial introgression occurring between both parental species and the Hawaiian duck at a whole-genome level (Table 1.2), as well as when analyzing chromosomes individually (Figure 1.1B; Supplementary Table S1.5). Similarly, calculating f_d statistics further supported varying admixture proportions

between the Hawaiian duck and its parental species. Notably, there is a higher degree of gene flow detected between Hawaiian and Laysan ducks at a whole-genome scale, but especially on the Z-sex chromosome (Table 1.3). Finally, I found small and interchanging blocks of Laysan duck and mallard ancestry across the Hawaiian duck's genome; although, there were more Laysan duck versus mallard ancestry blocks, in general (Figure 1.1C).

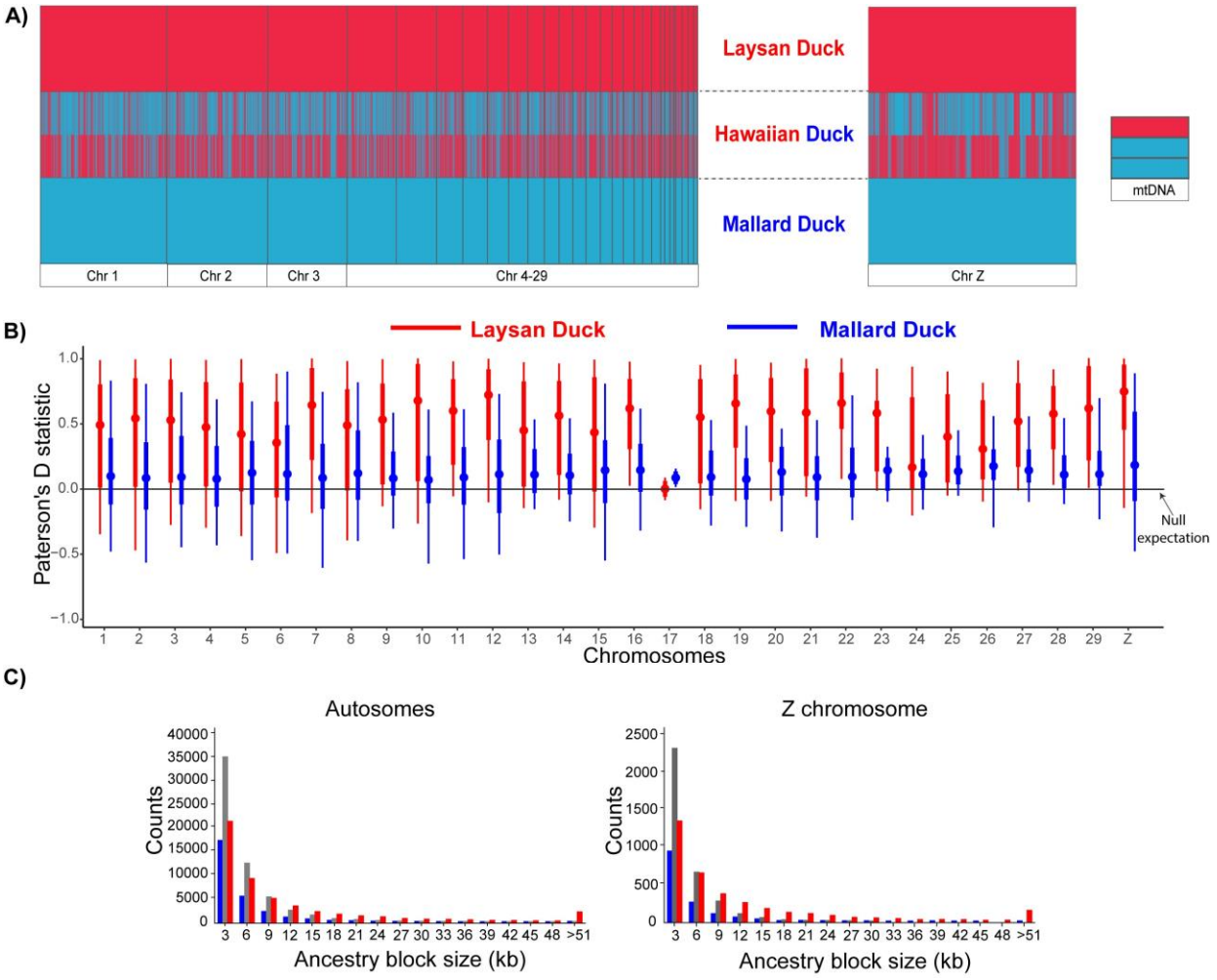

Figure 1.1. Parental contribution to the hybrid lineage. (A) Ancestry painting of Hawaiian duck's genome based fixed sites between Laysan duck and mallard, resulting in 624,665 base-pairs (bp) across autosomes (left), 44,007 bp across the Z-sex chromosome, and 15 bp for the mitochondrial DNA (mtDNA; right). (B) Per chromosome D-statistics within the Hawaiian duck's genome. (C) The size distribution of putative ancestry blocks from the Laysan duck (red), mallard (blue), and from unresolved ancestry (gray).

Table 1.2. Results from Patterson's D test for introgression between lineages with block Jack-knifed SE estimates and Z scores

	P1	P2	P3	Patterson's D	Jack-knifed SE	Z score
Autosomes	Laysan	Hawaiian	mallard	0.12	0.0040	30.42
	mallard	Hawaiian	Laysan	0.52	0.0080	66.44
Z-sex chromosome	Laysan	Hawaiian	mallard	0.17	0.019	8.78
	mallard	Hawaiian	Laysan	0.69	0.029	23.51

Table 1.3. Mean f_d values for 100-kb non-overlapping windows across the genomes

P1	P2	P3	f_d (%) Autosomes	f_d (%) Z chromosome
mallard	Laysan	Hawaiian	29.86 ± 27.61	38.14 ± 27.82
Laysan	mallard	Hawaiian	18.81 ± 9.26	3.32 ± 6.82

Phylogenetic relationships

Utilizing RAxML, I constructed maximum likelihood phylogenies across non-overlapping 100kb genomic windows, encompassing 267,102 autosomal SNPs and 21,045 Z-sex chromosome linked SNPs. Classifying whether the Hawaiian duck clustered monophyletically with the Laysan duck, the mallard or formed a distinct clade recovered a sister relationship between Hawaiian and Laysan ducks across the majority of trees (69.8%). However, in a minority of cases (17.8%), the Hawaiian duck clustered with the mallard, and even forming a unique clade in 12.4% of trees (Figure 1.2). A similar pattern emerged when examining Maximum Likelihood trees constructed across Z-sex chromosome windows where the Hawaiian and Laysan ducks were sister taxa in 86.2% of trees, with 7.5% and 6.4% of trees placing Hawaiian ducks as sister to mallards or forming a unique clade, respectively (Supplementary Figure S1.3).

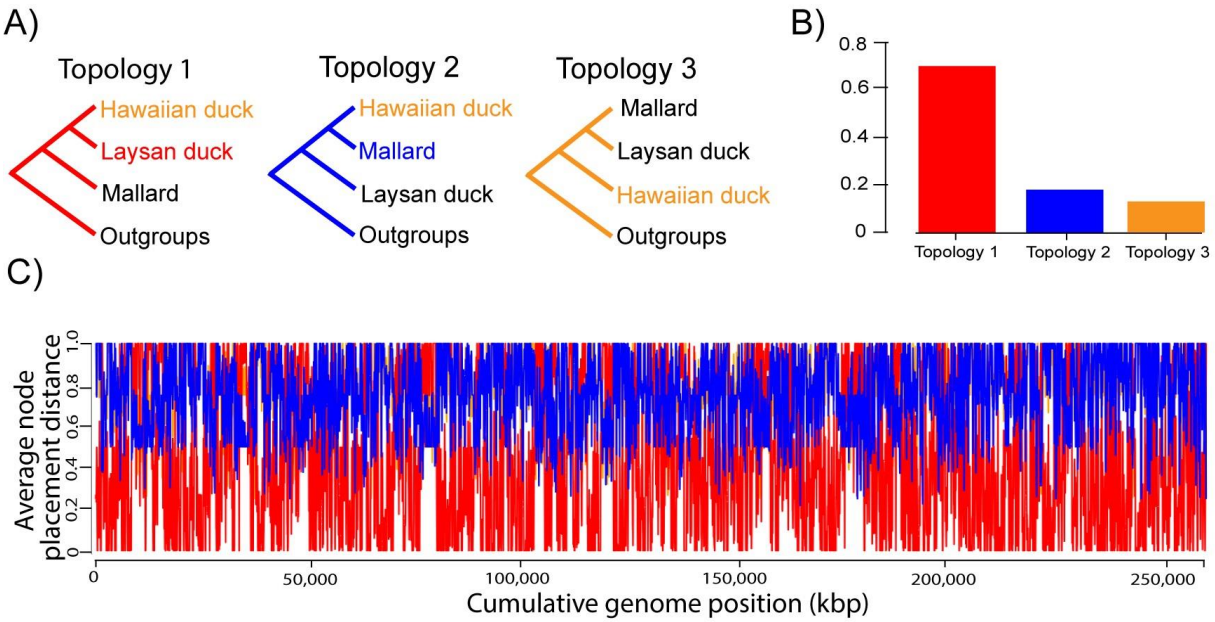

Figure 1.2. Phylogenetic inference across autosomes to assess the origin of the Hawaiian duck. (A) The main three tested topologies for the position of the Hawaiian duck, and (B) their relative weightings and (C) average node placement distance across autosomes.

Demographic histories

Reconstructing the effective ancestral population size (N_E) change across time revealed distinct demographic trajectories for mallards, Hawaiian ducks, and Laysan ducks (Figure 1.3). First, the effective population size of Laysan ducks diverged from Hawaiian ducks and mallards about a million years ago, when they experienced declines in effective population size, while Hawaiian ducks tracked mallards in an increasing trend. Interestingly, the Hawaiian duck diverged in pattern from the mallard ~570,000 years ago, at which point it began to mimic the demographic ancestry of the Laysan duck. The mallard exhibited a decline in its effective population size from its peak (N_E = 5 million), while the Laysan duck peaking at an estimated effective population size of 80,000 during the penultimate glacial period dating between 130,000 to 194,000 years ago. Finally, during the most recent ice age, the mallard maintained a stable

effective population of ~3 million, while both the Hawaiian and Laysan duck experienced significant declines.

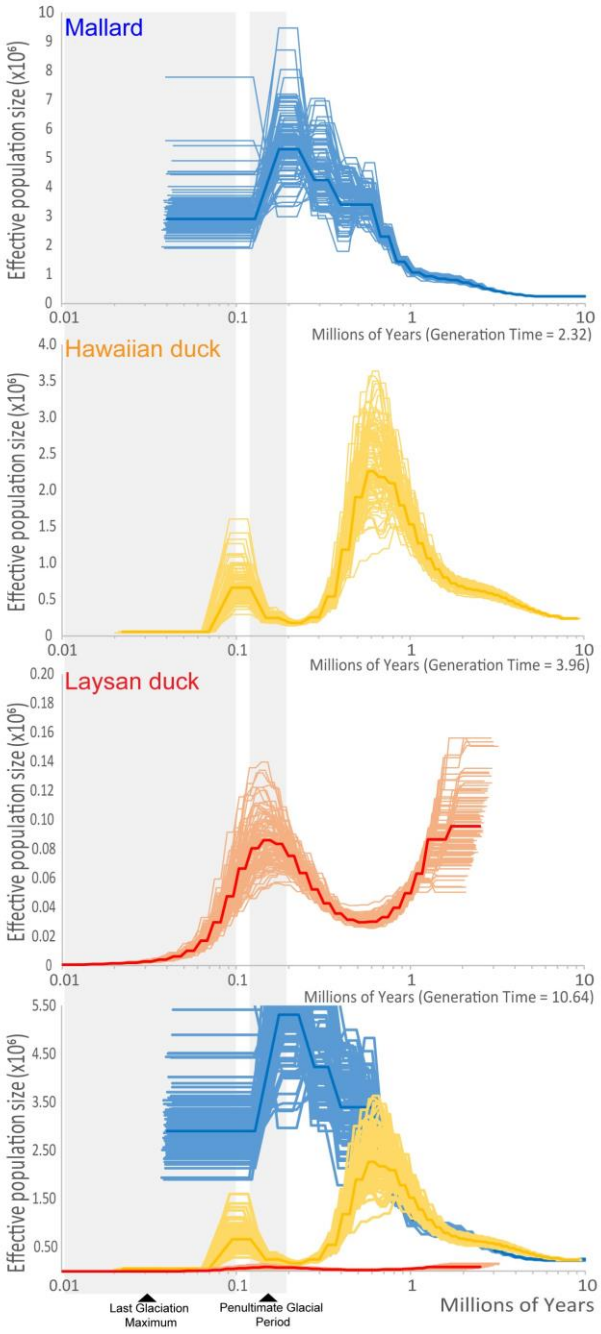

Figure 1.3. PSMC results. Demographic analyses of effective population size (N_E) calculated across the genomes of a wild mallard, Hawaiian duck, and Laysan duck. Note that the dark and lighter lines denote the average and bootstrap estimates, respectively.

Signatures of selection

Overall, I observed higher F_{ST} values, higher D_{XY} values, and lower nucleotide diversity within outlier windows compared to the non-outlier windows (Figure 1.4B-C). The Laysan-Hawaiian (LH) duck comparison exhibited the highest total number of outlier windows (97 in autosomes and 7 in the Z-sex chromosome), followed by an equal number of outlier windows in the mallard-Hawaiian (MH; 96 in autosomes and 7 in the Z-sex chromosome) and Hawaiian duck-both parent (PH) comparisons (96 in autosomes and 7 in the Z-sex chromosome).

Comparing F_{ST}, D_{XY}, and nucleotide diversity permitted us to associate outlier windows representing the "divergence-with-gene-flow" model that is characterized with high F_{ST}, high D_{XY}, and low nucleotide diversity (Irwin et al., 2018). Alternatively, outlier regions characterized by high F_{ST}, low D_{XY}, and low nucleotide diversity represented a model of "recurrent selection" or "geographic-sweep-before-selective-differentiation" (Irwin et al., 2018). A total number of windows that followed the "divergence-with-gene-flow" model included 104 windows in the PH comparison (Autosomes = 97, Z-sex chromosome = 8), 102 in the LH comparison (Autosomes = 94, Z-sex chromosome = 8), and 95 in the MH comparison (Autosomes = 89, Z-sex chromosome = 6). Outlier windows that were more consistent with an alternative model included 7 in the MH comparison (Autosomes = 6, Z-sex chromosome = 1), 3 for the LH comparison (Autosomes = 2, Z-sex chromosome = 1), and 1 autosomal window in the PH comparison (Figure 1.4). Finally, I recovered 3 outlier windows in the MH comparison (Autosomes = 2, Z-chromosome = 1) and 2 autosomal windows in the LH comparison that did not follow either model as I found high F_{ST}, high D_{XY}, and high nucleotide diversity; suggesting these regions may have experienced both divergence between populations and maintenance of diversity within populations.

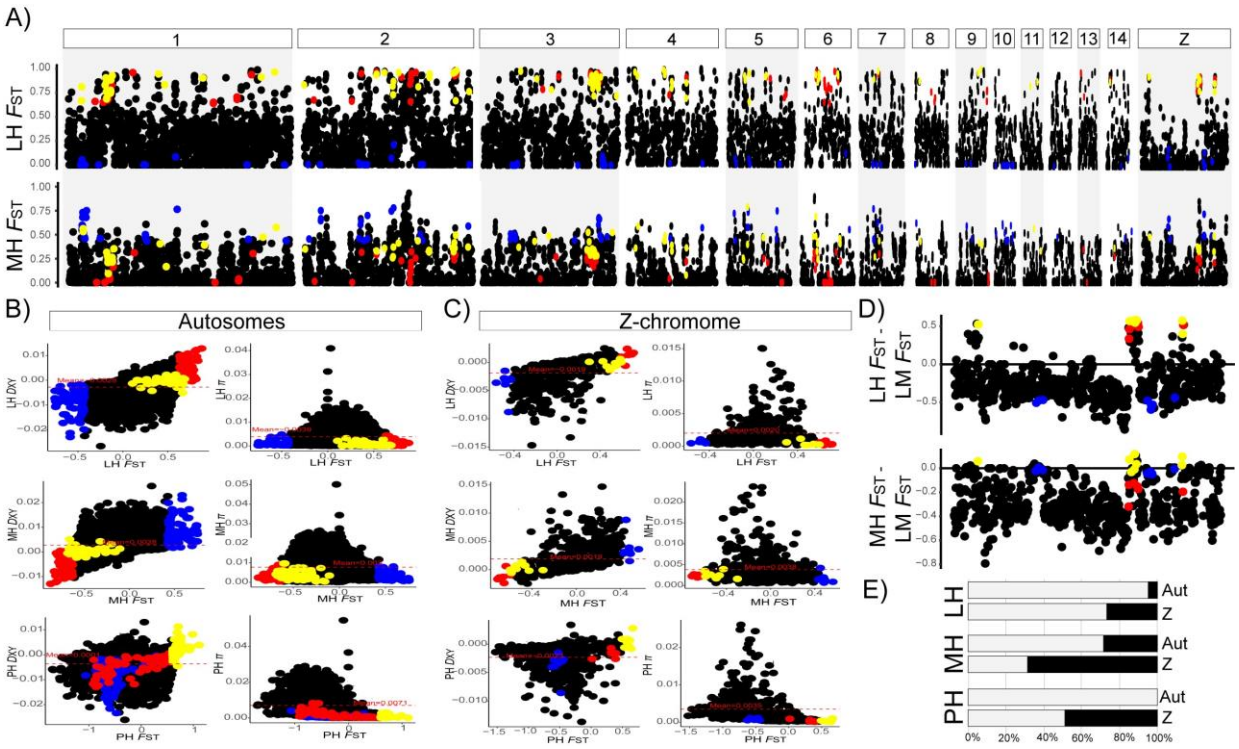

Figure 1.4. Divergence peaks between the Hawaiian duck and parent taxa. (A) F_{ST} estimates between Hawaiian ducks (H) and Laysan duck (L) (LH; top) or mallard (M) (MH; second from top) for 100kb non-overlapping windows across their genomes. (B) Correlation between F_{ST} and d_{XY} and between F_{ST} and nucleotide diversity in autosomes. (C) Correlation between F_{ST} and d_{XY} and between F_{ST} and nucleotide diversity in the Z-sex chromosome. (D) Disparities in F_{ST} values between each LH and MH comparison from the parent species comparison for the same 100kb non-overlapping windows in the Z-sex chromosome. Note that divergence windows recovered for LH, MH, and PH are color coded as red, blue, and yellow, respectively. Finally, (E) the distribution of outlier genes recovered across autosomes versus the Z-sex chromosome.

Outlier composition

Annotated genes residing within all the outlier windows were extracted for ontology analysis, resulting in a total of 217 LH, 186 MH, and 183 PH outlier genes (Supplementary Table S1.6, S1.7, S1.8, S1.9). Enrichment analysis revealed significant gene ontologies (GO) associated with processes that may contribute to species differentiation within the system. Notably, all significant GO terms were found in the outlier windows representing the "divergence-with-gene-flow" model, and many of these significant GO terms were found to be linked to the Z-sex

chromosome rather than autosomes (Figure 1.4E). These processes were grouped into seven distinct categories based on their function, including DNA information processing, cellular structure, metabolism and function, behavior, individual's well-being and health (e.g. sleep, appetite, mood, energy), neurotransmission, fertilization and embryonic development, and immune response. The predominant GO terms in the LH comparison (Supplementary Table S1.7) were associated with cellular structure metabolism and function, while eight GO terms were associated with DNA information processing. For MH comparison (Supplementary Table S1.8), I recovered a comparable number of GO terms covering various processes, including behavior, neurotransmission, individual's well-being and health, cellular structure, metabolism and function, and DNA information processing. A single GO term was not categorized because of lack of information. Finally, most GO terms were related to cellular structure, metabolism and function. Additionally, GO terms associated with facilitating fertilization and subsequent embryonic development were recovered in the PH comparison (Supplementary Table S1.9). Fewer GO terms were associated with individual's well-being and health, immune response, and behavior while one GO term was related to neurotransmission in the PH comparison.

Finally, KEGG pathway and enrichment analyses revealed only two significant pathways within all windows that potentially contribute to species differentiation (Supplementary Table S1.10, S1.11, S1.12). The "FoxO signaling" and the "ubiquitin-mediated proteolysis" pathways were identified on the Z-sex chromosome within the PH comparison. The FoxO signaling pathway plays a crucial role in environmental information processing and signal transduction (Eijkelenboom & Burgering, 2013; Farhan et al., 2017), while the ubiquitin-mediated proteolysis pathway is vital for genetic information processing, protein folding, sorting, and degradation

(Kipreos, 2005). It is important to highlight that several other KEGG pathways showed strong associations with the immune response, followed by connections to DNA transcription, as well as pathways associated with processes related to digestion, metabolism, and the circulatory system.

DISCUSSION

Genomic mosaicism and evolutionary dynamics of the Hawaiian duck

My study finds that the genome of the Hawaiian duck is indeed a mosaicism of Laysan duck and mallard genetic ancestry, and consistent with the hypothesis that they represent a young avian homoploid hybrid species (Lavretsky et al., 2015). Overall, average relative divergence estimates demonstrate a closer affinity between Hawaiian ducks and mallards, with my genomic means being like those estimated with a few loci (Table 1.1; Lavretsky et al., 2014, 2015). However, testing for phylogenetic relationships across the genome placed Laysan and Hawaiian ducks as sister taxa in majority of analyzed trees regardless of chromosome-type (Figure 1.2). The disparity between general F_{ST} estimates and phylogenetic relatedness is likely because of increased homozygosity across the Laysan duck's genome (Table 1.1) as a result of them going through severe bottlenecking to a population size of 11 individuals by 1911 (M. Reynolds & Klavitter, 2006). Nevertheless, whole genome admixture is supported in my ABBA-BABA (Figure 1.1B; Table 1.2) and f_d (Table 1.3) estimates, while more refined ancestry analyses recovered an interchange of relatively small blocks (i.e., ~ 3 Kbp; Figure 1.1C) of Laysan duck and mallard ancestry across the Hawaiian duck's genome (Figure 1.1A). Together, these results are consistent with an ancient introgression event followed by backcrossing (Meier et al., 2017; Runemark et al., 2018). Despite an overall admixed genome, 6-12% of phylogenetic trees placed Hawaiian ducks as a unique clade (Figure 1.2), and with divergence calculations recovering a

similar proportion of loci as unique to Hawaiian ducks (Figure 1.4). This suggests that divergent selection has already played a role in shaping the Hawaiian duck's genome since their hybrid beginnings. I conclude that the Hawaiian duck represents an ancient hybrid lineage that is gradually differentiating from their parental taxa (Schumer et al., 2014).

Often, the evolution of a hybrid lineage occurs either via equal genomic contribution of parental species or through the adaptive introgression of a few genes from one parental species to the genome of the other (R. Abbott et al., 2013a; R. J. Abbott et al., 2010; Jiggins et al., 2008; Schumer et al., 2014). For the Hawaiian duck, their genomic mosaicism suggests that both parental taxa substantially contributed to their genome, and thus they do not represent a case of adaptive introgression. Nevertheless, the slight bias in ancestry between Hawaiian and Laysan ducks may be a result of backcrossing at initial stages of hybridization (Elgvin et al., 2017) as Laysan ducks very likely outnumbered vagrant mallards at the time (Lavretsky et al., 2015), and thus, resulting in an overrepresentation of the genetic background of the former species. Alternatively, the observed bias could be a result of Laysan ducks harboring allelic diversity adapted to Island life, and thus, those similar selective pressures also exhibited on these early hybrids to retain the same genetic diversity. Unlike the nuclear genome, I can confirm that Hawaiian ducks retained the mallard mitogenome (Lavretsky et al., 2014, 2015). Whether this capture of the mallard mitogenome in the Hawaiian duck's lineage was adaptive or stochastic remains unknown. However, given that mallards were the non-native parental taxa in this case, retention of its mitogenome because of local adaptive qualities is unlikely. Thus, I hypothesize that this mitogenome capture event was likely because of more random effects of genetic drift. Using ancient DNA techniques to sequence historical and subfossil remains of putative Hawaiian ducks found across biodiversity collections can help further test between these scenarios.

Towards this end, this study provides the first reference genomes for the endangered Laysan and Hawaiian ducks that can serve as references for such ancient DNA work, as well as important tools in the future of their conservation.

Ancestral population size and hybrid genome influences

The analysis of effective ancestral population size (N_E) reveals distinct demographic trajectories among mallards, Hawaiian ducks, and Laysan ducks (Figure 1.3), influenced by colonization events, founder effects, impact of glaciation, and a hybrid genome. Divergence time for Laysan ducks is consistent with previous speciation times for the Mallard Complex more generally, dating to a million years ago when an African ancestral mallard-like duck began to simultaneously expand northward and through the South Pacific (Lavretsky et al., 2014). Notably, Laysan ducks exhibit the smallest effective population size since the beginning of their speciation, which combined with their decreased nucleotide diversity and extended runs of homozygosity, underscores a genetic history marked by isolation. During the penultimate glacial period, while mallards experienced a population decline, Laysan ducks reached a peak in their effective population size. These contrasting patterns can be attributed to the different impacts of the glaciation period, which particularly affected Eurasia, where the mallard resides, while potentially exerting less influence on the Hawaiian Islands where Laysan ducks evolved. However, during the last glaciation maximum, mallards established an effective population size of ~3 million, whereas Laysan ducks continued their decline. This extended period of glaciation likely affected Laysan duck populations in the Hawaiian Islands, while mallards had already expanded to North America and found refuges. Generally, however, the demographic history of the Hawaiian duck is a combination of mallard and Laysan duck histories and consistent with a

history of an admixed genome. Consequently, the reliability of Hawaiian duck divergence times in this analysis is compromised, as hybridization or extreme genetic drift events are recognized factors that can significantly distort such estimates (Brown et al., 2022; Lavretsky et al., 2023a; Orozco-terwengel & Bruford, 2014).

Insights into Hawaiian duck evolution and adaptation

Outlier windows identified through the comparison of the Hawaiian duck with each parent (LH for Laysan-Hawaiian and MH for mallard-Hawaiian) revealed key genetic variations for the Hawaiian duck's development of ecological, behavioral, and physiological distinctions. As a result, these identified regions emerge as potential candidates for divergence barriers. First, key GO terms in the outlier windows of the Laysan duck versus Hawaiian duck comparison shed light on physiological processes (Supplementary Table S1.7). Specifically, GO terms associated with cholesterol and isoprenoid biosynthesis, along with coenzyme A metabolic processes, indicate potential influences on the physiological characteristics such as lipid metabolism (Brockmeier et al., 2016; Lee et al., 2012). Additionally, the comparison between the mallard and Hawaiian duck also revealed a spectrum of GO terms representing diverse biological processes and molecular functions (Supplementary Table S1.8). Findings include neurotransmitter-related terms (e.g., serotonin binding, G protein-coupled serotonin receptor activity) and processes regulating serotonin secretion and vasoconstriction. These indicate distinctions in neurotransmitter systems and various physiological responses, spanning areas such as appetite and digestion (Meguid et al., 2000), blood pressure (Kour et al., 2023), olfactory receptors (Grapputo et al., 2018; Schrader et al., 2014), and behavioral regulation (Cantabella et al., 2022; Scott et al., 2022).

The significance of these GO terms lies in their potential contribution to the Hawaiian duck's adaptation to a distinct ecological niche compared to its parent species. For instance, the restricted distribution of Laysan ducks results in these exclusively feeding on macroinvertebrates (Moulton & Marshall, 2020), while the mallard can occur in almost every wetland and has an opportunistic and omnivorous diet (Drilling et al., 2020). The Hawaiian duck exhibits intermediate ecological traits, living in a wider range of habitats as compared to Laysan ducks, and having a broader diet, including aquatic invertebrates, aquatic plants, seeds, and grains (Engilis Jr. et al., 2020a). Similarly, notable findings in the mallard and Hawaiian duck comparison include behavioral GO terms such as adult behavior, behavioral fear response, and exploration behavior hint at unique behavioral traits in each species, possibly reflecting ecological adaptations or social behaviors (Supplementary Table S1.8). The significance of these behaviors becomes apparent when considering the survival of the Hawaiian duck, as it had to cope with predators on the Hawaiian Islands, including threats during human colonization (e.g. habitat destruction, introduced species, overhunting; Athens et al., 2002; Boyer, 2008). I posit that having such traits likely explains how the evolving hybrid swarm of Hawaiian ducks evaded extinction, while the Laysan duck did not (M. H. Reynolds et al., 2015). Thus, evolution and adaptive capacity of the Hawaiian duck, particularly during Polynesian colonization may be related to social behaviors more akin to the mallard than the Laysan duck. Overall, the Hawaiian duck’s admixed genome may explain the diversity of functions the species displays, shedding light into how they evolved such unique ecological, behavioral, and physiological features.

Finally, the comparative analysis between the Hawaiian duck and its parental species (PH) reveals distinct genomic regions associated with immune response and reproductive

isolation, marking potential adaptations crucial for the survival of this emerging hybrid lineage (Supplementary Table S1.8). Notably, key GO terms like "azurophil granule" (GO:0042582) assume significance in the immune system, suggesting a potential role in the Hawaiian duck's resilience against malaria, a pressing concern for many native bird species in Hawaii (Beadell et al., 2006). Furthermore, the identification of significant gene ontologies such as 'male courtship behavior' and 'in utero embryonic development,' exclusive to the Z-sex chromosome holds particular importance in the context of reproductive isolation. The presence of these significant GO terms is consistent with the role sex chromosomes play in shaping reproductive barriers and facilitating the emergence of distinct species (Peters et al., 2005; Sætre & Sæther, 2010). These GO terms suggest that the Hawaiian duck is likely proceeding through initial phases of species divergence, including likely build-up of reproductive isolation. It is crucial to acknowledge that the formation of a complete reproductive barrier is a gradual process, often spanning millions of years in birds (Gill, 2014; Grant & Grant, 1992a; Lavretsky et al., 2015). This protracted timeline indicates that the Hawaiian duck represents a relatively recent instance of hybrid speciation, but emphasizing the intricate and dynamic nature of the evolutionary processes leading to the emergence of novel lineages.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., & Brennan, A. C. (2010). Homoploid hybrid speciation in action. *TAXON*, *59*(5), 1375–1386. https://doi.org/10.1002/tax.595005

Alerstam, T., & Högstedt, G. (1982). Bird Migration and Reproduction in Relation to Habitats for Survival and Breeding. *Ornis Scandinavica (Scandinavian Journal of Ornithology)*, *13*(1), 25–37. https://doi.org/10.2307/3675970

Aliabadian, M., & Nijman, V. (2007). Avian hybrids: Incidence and geographic distribution of hybridisation in birds. *Contributions to Zoology*, *76*(1), 59–61. https://doi.org/10.1163/18759866-07601006

Arnold, T. W., & Clark, R. G. (1996). Survival and Philopatry of Female Dabbling Ducks in Southcentral Saskatchewan. *The Journal of Wildlife Management*, *60*(3), 560–568. https://doi.org/10.2307/3802073

Athens, J. S., Toggle, H. D., Ward, J. V., & Welch, D. J. (2002). Avifaunal extinctions, vegetation change, and Polynesian impacts in prehistoric Hawai'i. *Archaeology in Oceania*, *37*(2), 57–78. https://doi.org/10.1002/j.1834-4453.2002.tb00507.x

Barth, J. M. I., Gubili, C., Matschiner, M., Tørresen, O. K., Watanabe, S., Egger, B., Han, Y.-S., Feunteun, E., Sommaruga, R., Jehle, R., & Schabetsberger, R. (2020). Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15099-x

Barton, N. H. (2013). Does hybridization influence speciation? *Journal of Evolutionary Biology*, *26*(2), 267–269. https://doi.org/10.1111/jeb.12015

Beadell, J. S., Ishtiaq, F., Covas, R., Melo, M., Warren, B. H., Atkinson, C. T., Bensch, S., Graves, G. R., Jhala, Y. V., Peirce, M. A., Rahmani, A. R., Fonseca, D. M., & Fleischer, R. C. (2006). Global phylogeographic limits of Hawaii's avian malaria. *Proceedings of*

the Royal Society B: Biological Sciences, *273*(1604), 2935–2944. https://doi.org/10.1098/rspb.2006.3671

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boyer, A. G. (2008). Extinction patterns in the avifauna of the Hawaiian islands. *Diversity and Distributions*, *14*(3), 509–517. https://doi.org/10.1111/j.1472-4642.2007.00459.x

Brockmeier, E. K., Scott, P. D., Denslow, N. D., & Leusch, F. D. L. (2016). Transcriptomic and physiological changes in Eastern Mosquitofish (*Gambusia holbrooki*) after exposure to progestins and anti-progestagens. *Aquatic Toxicology*, *179*, 8–17. https://doi.org/10.1016/j.aquatox.2016.08.002

Brown, J. I., Harrigan, R. J., & Lavretsky, P. (2022). Evolutionary and ecological drivers of local adaptation and speciation in a North American avian species complex. *Molecular Ecology*, *31*(9), 2578–2593. https://doi.org/10.1111/mec.16423

Cantabella, E., Camilleri, V., Cavalie, I., Dubourg, N., Gagnaire, B., Charlier, T. D., Adam-Guillermin, C., Cousin, X., & Armant, O. (2022). Revealing the Increased Stress Response Behavior through Transcriptomic Analysis of Adult Zebrafish Brain after Chronic Low to Moderate Dose Rates of Ionizing Radiation. *Cancers*, *14*(15), Article 15. https://doi.org/10.3390/cancers14153793

Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2020). *Genome-wide local ancestries discriminate homoploid hybrid speciation from secondary introgression in the red wolf (Canidae: Canis rufus)* (p. 2020.04.05.026716). bioRxiv. https://doi.org/10.1101/2020.04.05.026716

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Drilling, N., Titman, R. D., & McKinney, F. (2020). Mallard (Anas platyrhynchos), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.mallar3.01species_shared.bow.project_name

Durand, E. Y., Patterson, N., Reich, D., & Montgomery, S. (2011). *Testing for Ancient Admixture between Closely Related Populations*. *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Eijkelenboom, A., & Burgering, B. M. T. (2013). FOXOs: Signalling integrators for homeostasis maintenance. *Nature Reviews Molecular Cell Biology*, 14(2), 83–97. https://doi.org/10.1038/nrm3507

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Engilis Jr., A., Uyehara, K. J., & Giffin, J. G. (2020). Hawaiian Duck (Anas wyvilliana), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.hawduc.01species_shared.bow.project_name

Farhan, M., Wang, H., Gaur, U., Little, P. J., Xu, J., & Zheng, W. (2017). FOXO Signaling Pathways as Therapeutic Targets in Cancer. *International Journal of Biological Sciences*, *13*(7), 815–827. https://doi.org/10.7150/ijbs.20052

Gill, F. B. (2014). *Species taxonomy of birds: Which null hypothesis?Taxonomía de especies de aves: ¿Cuál es la hipótesis nula?Species taxonomy of birds | Ornithology | Oxford Academic*. *131*(2), 150–161.

Grant, P. R., & Grant, B. R. (1992). Hybridization of bird species. *Science (New York, N.Y.)*, *256*(5054), 193–197. https://doi.org/10.1126/science.256.5054.193

Grapputo, A., Thrimawithana, A. H., Steinwender, B., & Newcomb, R. D. (2018). Differential gene expression in the evolution of sex pheromone communication in New Zealand's endemic leafroller moths of the genera Ctenopseustis and Planotortrix. *BMC Genomics*, *19*(1), 94. https://doi.org/10.1186/s12864-018-4451-1

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science (New York, N.Y.)*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hoff, K. J., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*(1), e57. https://doi.org/10.1002/cpbi.57

Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, *27*(23), 4839–4855. https://doi.org/10.1111/mec.14858

Jiggins, C. D., Salazar, C., Linares, M., & Mavarez, J. (2008). Hybrid trait speciation and Heliconius butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3047–3054. https://doi.org/10.1098/rstb.2008.0065

Johnsgard, P. A. (1960). Hybridization in the Anatidae and Its Taxonomic Implications. *The Condor*, *62*(1), 25–33. https://doi.org/10.2307/1365656

Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, *30*(1), 42–46. https://doi.org/10.1093/nar/30.1.42

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kipreos, E. T. (2005). Ubiquitin-mediated pathways in C.elegans. In *WormBook: The Online Review of C. elegans Biology [Internet]*. WormBook. https://www.ncbi.nlm.nih.gov/books/NBK19761/

Kour, A., Deb, S. M., Nayee, N., Niranjan, S. K., Raina, V. S., Mukherjee, A., Gupta, I. D., & Patil, C. S. (2023). Novel insights into genome-wide associations in Bos indicus reveal genetic linkages between fertility and growth. *Animal Biotechnology*. https://www.tandfonline.com/doi/abs/10.1080/10495398.2021.1932520

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science (New York, N.Y.)*, *359*(6372), 224–228. https://doi.org/10.1126/science.aao4593

Lavretsky, P. (2021). Population Genomics Provides Key Insights into Admixture, Speciation, and Evolution of Closely Related Ducks of the Mallard Complex. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population Genomics: Wildlife* (pp. 295–330). Springer International Publishing. https://doi.org/10.1007/13836_2020_76

Lavretsky, P., Engilis Jr, A., Eadie, J. M., & Peters, J. L. (2015). Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. *Journal of Evolutionary Biology*, *28*(5), 1005–1015. https://doi.org/10.1111/jeb.12637

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. (2023a). *Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos)*. *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. E. (2023b). Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos). *G3 Genes|Genomes|Genetics*, *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., McCracken, K. G., & Peters, J. L. (2014). Phylogenetics of a recent radiation in the mallards and allies (Aves: Anas): Inferences from a genomic transect and the multispecies coalescent. *Molecular Phylogenetics and Evolution*, *70*, 402–411. https://doi.org/10.1016/j.ympev.2013.08.008

Lavretsky, P., McInerney, N. R., Mohl, J. E., Brown, J. I., James, H. F., McCracken, K. G., & Fleischer, R. C. (2020). Assessing changes in genomic divergence following a century of

human-mediated secondary contact among wild and captive-bred ducks. *Molecular Ecology*, *29*(3), 578–595. https://doi.org/10.1111/mec.15343

Lee, S. Y., Sohn, K.-A., & Kim, J. H. (2012). MicroRNA-centric measurement improves functional enrichment analysis of co-expressed and differentially expressed microRNA clusters. *BMC Genomics*, *13*(7), S17. https://doi.org/10.1186/1471-2164-13-S7-S17

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), Article 7357. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Livezey, B. C. (1993). Comparative morphometrics of Anas ducks, with particular reference to the Hawaiian Duck Anas wyvilliana, Laysan Duck A. laysanensis, and Eaton's Pintail A. eatoni. *Wildfowl*, *44*(44), Article 44.

Malachowski, C. P., Dugger, B. D., Uyehara, K. J., & Reynolds, M. H. (2022). Avian botulism is a primary, year-round threat to adult survival in the endangered Hawaiian Duck on Kaua'i, Hawai'i, USA. *Ornithological Applications*, *124*(2), duac007. https://doi.org/10.1093/ornithapp/duac007

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), Article 7133. https://doi.org/10.1038/nature05706

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature*, *441*(7095), 868–871. https://doi.org/10.1038/nature04738

Meguid, M. M., Fetissov, S. O., Blaha, V., & Yang, Z. J. (2000). Dopamine and serotonin VMN release is related to feeding status in obese and lean Zucker rats. *Neuroreport*, *11*(10), 2069–2072. https://doi.org/10.1097/00001756-200007140-00002

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14363

Moulton, D. W., & Marshall, A. P. (2020). Laysan Duck (Anas laysanensis), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.layduc.01species_shared.bow.project_name

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology: CB*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Nolte, A. W., & Tautz, D. (2010). Understanding the onset of hybrid speciation. *Trends in Genetics*, *26*(2), 54–58. https://doi.org/10.1016/j.tig.2009.12.001

Orozco-terwengel, P. A., & Bruford, M. W. (2014). Mixed signals from hybrid genomes. *Molecular Ecology*, *23*(16), 3941–3943. https://doi.org/10.1111/mec.12863

Ortiz, E. M. (2019). *vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis* [DOI:10.5281/zenodo.2540861]. https://zenodo.org/records/2540861/preview/edgardomortiz/vcf2phylip-v2.0.zip?include_deleted=0

Peters, J. L., McCracken, K. G., Zhuravlev, Y. N., Lu, Y., Wilson, R. E., Johnson, K. P., & Omland, K. E. (2005). Phylogenetics of wigeons and allies (Anatidae: Anas): the importance of sampling multiple loci and multiple individuals. *Molecular Phylogenetics and Evolution*, *35*(1), 209–224. https://doi.org/10.1016/j.ympev.2004.12.017

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*(3), 342–350. https://doi.org/10.1101/gr.193474.115

Reynolds, M. H., & Citta, J. J. (2007). Postfledging Survival of Laysan Ducks. *The Journal of Wildlife Management*, *71*(2), 383–388.

Reynolds, M. H., Pearce, J. M., Lavretsky, P., L, P. J., Courtot, K., & Seixas, P. P. (2015). Evidence of low genetic variation and rare alleles in a bottlenecked endangered island endemic, the Lasan Teal (Anas laysanensis). In *Technical Report* (HCSU-063; pp. 1–14). University of Hawaii at Hilo. https://pubs.usgs.gov/publication/70144555

Reynolds, M., & Klavitter, J. (2006). Translocation of wild Laysan duck Anas laysanenis to establish a population at Midway Atoll National Wildlife Refuge, United States and US Pacific Possession. *Conservation Evidence*, *3*, 6–8.

Rheindt, F. E., & Edwards, S. V. (2011). Genetic Introgression: An Integral but Neglected Component of Speciation in Birds. *The Auk*, *128*(4), 620–632. https://doi.org/10.1525/auk.2011.128.4.620

Rhymer, J. (2006). *S33-4 Extinction by hybridization and introgression in anatine ducks*. *52*(Supplement), 583–585.

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), Article 4. https://doi.org/10.1038/sj.hdy.6886170

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science (New York, N.Y.)*, *317*(5840), 910–914. https://doi.org/10.1126/science.1137729

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., Elgvin, T. O., & Sætre, G.-P. (2018). Variation and constraints in hybrid genome formation. *Nature Ecology & Evolution*, *2*(3), Article 3. https://doi.org/10.1038/s41559-017-0437-7

Sæther, B.-E., Engen, S., Møller, A. P., Visser, M. E., Matthysen, E., Fiedler, W., Lambrechts, M. M., Becker, P. H., Brommer, J. E., Dickinson, J., du Feu, C., Gehlbach, F. R., Merilä, J., Rendell, W., Robertson, R. J., Thomson, D., & Török, J. (2005). Time to Extinction of Bird Populations. *Ecology*, *86*(3), 693–700.

Sætre, G.-P., & Sæther, S. A. (2010). Ecology and genetics of speciation in Ficedula flycatchers. *Molecular Ecology*, *19*(6), 1091–1106. https://doi.org/10.1111/j.1365-294X.2010.04568.x

Salazar, C. A., Jiggins, C. D., Arias, C. F., Tobler, A., Bermingham, E., & Linares, M. (2005). Hybrid incompatibility is consistent with a hybrid origin of Heliconius heurippa Hewitson from its close relatives, Heliconius cydno Doubleday and Heliconius

melpomene Linnaeus. *Journal of Evolutionary Biology*, *18*(2), 247–256. https://doi.org/10.1111/j.1420-9101.2004.00839.x

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, *5*(1), 5495. https://doi.org/10.1038/ncomms6495

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How Common Is Homoploid Hybrid Speciation? *Evolution*, *68*(6), 1553–1560. https://doi.org/10.1111/evo.12399

Schwarz, D., Matta, B. M., Shakir-Botteri, N. L., & McPheron, B. A. (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature*, *436*(7050), Article 7050. https://doi.org/10.1038/nature03800

Scott, A. M., Yan, J. L., Baxter, C. M., Dworkin, I., & Dukas, R. (2022). The genetic basis of variation in sexual aggression: Evolution versus social plasticity. *Molecular Ecology*, *31*(10), 2865–2881. https://doi.org/10.1111/mec.16437

Smith, G. W., & Reynolds, R. E. (1992). Hunting and Mallard Survival, 1979-88. *The Journal of Wildlife Management*, *56*(2), 306–316. https://doi.org/10.2307/3808827

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Uyehara, K. J., Engilis Jr., A., & Reynolds, M. (2007). *Hawaiian Duck's Future Threatened by Feral Mallards* (Fact Sheet 3047; Fact Sheet, pp. 1–4). U.S. Fish and Wildlife Service.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(W1), W52–W58. https://doi.org/10.1093/nar/gkz333

Chapter 2: simRestore: A decision-making tool for adaptive management of the native genetic status of wild populations

ABSTRACT

Anthropogenic hybridization, or higher and non-natural rates of gene flow directly and indirectly induced by human activities, is considered a significant threat to biodiversity. The primary concern for conservation is the potential for genomic extinction and loss of adaptiveness for native species because of the extensive introgression of non-native genes. To alleviate or reverse trends for such scenarios requires the direct integration of genomic data within a model framework for effective management. Towards this end, I developed the simRestore R program as a decision-making tool that integrates ecological and genomic information to simulate ancestry outcomes from optimized conservation strategies. The program optimizes supplementation and removal strategies across generations until a set native genetic threshold is reached within the studied population. Importantly, in addition to helping with initial decision-making, simulations can be updated with the outcomes of ongoing efforts, allowing for the adaptive management of populations. After demonstrating functionality, I apply and optimize among actionable management strategies for the endangered Hawaiian duck for which the current primary threat is genetic extinction through ongoing anthropogenic hybridization with feral mallards. Simulations demonstrate that supplemental and removal efforts can be strategically tailored to move the genetic ancestry of Hawaii's hybrid populations towards Hawaiian duck without the need to completely start over. Further, I discuss ecological parameter sensitivity, including which factors are most important to ensure genetic outcomes (i.e. number of offspring). Finally, to facilitate use, the program is also available online as a Shiny Web application.

INTRODUCTION

Classically, gene flow or the interbreeding of individuals from genetically distinct taxonomic units is often expected to result in outbreeding depression (R. Abbott et al., 2013b; Barton, 2013b; Mallet, 2007b), with sustained exchange of genetic material eventually leading to the loss of either one (i.e., loss of taxa) or both (i.e., formation of a hybrid swarm) interacting species (Kearns et al., 2018; Seehausen, Takimoto, Roy, & Jokela, 2008). However, secondary contact between taxa can also lead to the formation and maintenance of hybrid zones, which can cause reproductive reinforcement, and perhaps eventual completion of the speciation process (R. Abbott et al., 2013b; Grant & Grant, 1992b; López-Caamal & Tovar-Sánchez, 2014; J. M. Rhymer, 2006; Todesco et al., 2016). While gene flow is now recognized as an important natural phenomenon in the evolutionary history of many animal species, human activities have dramatically increased the rates of hybridization worldwide (Allendorf et al., 2001; Mallet, 2005b; Nolte & Tautz, 2010b). Specifically, direct and indirect human activities (e.g., increased urbanization, augmenting wild lands, and the intentional and unintentional release of invasive and often domestic species) are leading to unnaturally high rates of secondary contact among historically allopatric species (Crispo et al., 2011; McFarlane & Pemberton, 2019). Such anthropogenic hybridization has become a focal cause of concern for the conservation of many species (Crispo et al., 2011; Leitwein, Gagnaire, Desmarais, Berrebi, & Guinand, 2018; McFarlane & Pemberton, 2019; Wells et al. 2019; Lavretsky et al. 2023). Among activities that result in such human-mediated gene flow is the common practice of using captive-reared populations for conservation or restocking purposes in forestry, fisheries, and game management (Brennan et al., 2014; Söderquist et al., 2017). However, the selective pressures on wild (i.e., natural selection) versus domestic (i.e., artificial selection) individuals often results in contrasting

trait selection, with those specific to human-modified environments often being maladaptive in the wild (Christie et al., 2012;Crispo et al., 2011). Thus, the interbreeding between domestic and wild counterparts has frequently been found to lead to outbreeding depression or reduced local adaptation in the wild (Crispo et al., 2011). Increasing incidence of such interactions has brought understanding the impacts of anthropogenic hybridization on wild populations to the forefront of conservation science (Hirashiki et al., 2021).

Gene flow is both perceived as a problem and heralded as a potential solution, depending on the taxonomic organism of interest within conservation science (Flanagan et al., 2018). On the one hand, a species under threat of genetic extinction can require management involving the removal of the invading species, translocation of the threatened population, or habitat improvement (Rieseberg & Gerber, 1995; Wolf et al., 2001). Conversely, gene flow has been used to a limited extent as a conservation strategy to rescue the viability (i.e., improve the fitness) of small, inbred populations, known as 'genetic rescue' (Frankham, 2015; Hedrick & Fredrickson, 2010; Miller et al., 2012; Todesco et al., 2016). Traditional methods for evaluating these conservation actions have been based on increases in positive (e.g., population size, reproductive success, and survival rates) and decreases in negative (e.g., deleterious traits and mortality rate) ecological factors of the species to be conserved(Frankham, 2015; Hedrick & Fredrickson, 2010; Miller et al., 2012). Advances in next-generation DNA sequencing technologies along with novel analytical methods have been useful to provide genetic information, such as rates of hybridization, to complement conservation plans (Anderson & Thompson, 2002; Flanagan et al., 2018; Hohenlohe et al., 2021; van Wyk et al., 2017). However, integrating genomic data into conservation management requires significant genetic knowledge

and bioinformatics expertise, which is often lacking (Flanagan et al., 2018; Hoban et al., 2013; Hohenlohe et al., 2021).

To reach policymakers and managers, the development of more user-friendly programs and clear guidelines for applying genetic information to wildlife biology and management is needed. In particular, methodologies in which molecular and/or ecological data can be annually updated would provide a means for adaptive management planning. Towards this end, I developed simRestore, a decision-making R-based program that simulates the time (in generations) until a population may attain genetic native status under differing management strategies. The program makes use of backcrossing and admixture as a mechanism to establish genetic integrity (Lavretsky et al., 2016, 2019), and provides users a framework to incorporate management actions (e.g., augmenting and removal efforts) in combination with ecologically informative variables (e.g., survival rate) and genetic information (e.g., ancestry assignment) to simulate the expected time in generations to achieve the native status of the threatened species. Thus, the program consists of two intertwined models covering ecological and genetic data (Table 2.1). Importantly, as most conservation programs are resource-limited, the program is also designed to simulate under those limitations, including the total number of individuals that can be released and/or removed, as well as any temporal restrains for project completion. Finally, in addition to helping with initial decision making, simulations can be updated with the outcomes of ongoing efforts, allowing for the adaptive management of populations. Here, I demonstrate software functionality, as well as its utility through optimization of actionable management strategies for the endangered Hawaiian duck (*Anas wyvilliana*) for which the current primary threat is genetic extinction through ongoing anthropogenic hybridization with feral mallards (*Anas platyrhynchos*) (USFWS, 2012; Wells et al., 2019).

Study system

Once found across the main Hawaiian Islands, Hawaiian ducks once again only reside on the Island of Kauai (Wells et al., 2019). Although habitat loss and overhunting were in part responsible for their decline (Engilis & Pratt, 1993), it is the establishment of feral mallard populations that appear to be the proximate cause of conservation concern today (Engilis Jr., Uyehara, & Giffin, 2020, see also Wells et al. 2019). Although captive-rearing programs and reintroductions were attempted from the1960s to the 1980s (Browne et al., 1993; Engilis & Pratt, 1993), these efforts ultimately failed because of not handling the burgeoning feral mallard populations that eventually interbred with translocated Hawaiian ducks to form feral mallard x Hawaiian duck hybrid populations across Islands (Wells et al., 2019). Although the conservation of the Hawaiian duck could move forward with restarting all populations from Kauai stock after the extirpation of existing feral mallard x Hawaiian duck hybrids, such efforts are often not financially sustainable but also are complicated by human-dimensions (Stronen & Paquet, 2013). Instead, there is potential to maximize management strategies by varying restocking and partial removal efforts optimized for each wetland's characteristics. Given that sequential backcrossing into the same gene pool has the potential to re-establish the genetic signature of the backcrossed parental population within only a few generations (Lavretsky et al., 2016, 2019), the genetics of a hybrid population can thus be artificially moved towards a target parental through directed management efforts. However, the number of individuals required to be added and/or removed is dependent on many ecological factors (see Table 2.1; Hennessy et al., 2022). Fortunately, continued conservation efforts surrounding Hawaiian ducks has resulted in in-depth biological and genetic knowledge for the species that can be used to optimize management strategies among populations.

Table 2.1. Overview of available parameters available in the simRestore R package as well as their range of values to modify depending on the studies species. All values used across functions for Hawaiian duck simulations are also provided.

Function	Function Description	Range of Values	Hawaiian Duck Simulation Input Values
Morgan	Size of the chromosome in Morgan	0+	1 (Lavretsky et al. 2019)
K	Carrying capacity	1+	400 (Robinson et al. 2017)
Reproductive success rate	frequency of females that yield offspring at the end of the breeding season (e.g. a fraction of 1 - reproduction_success_rate of females. This is a joint effect of breeding females getting killed (see female_death_rate) and other sources of failure to complete a clutch. Other sources of failure are calculated from nest_success_rate and female_death_rate, such that reproduction failure rate = 1 - reproduction_success_rate / (1 - female breeding risk);	From 0 to 1	0.387 (Malachowski et al. 2018)
Reproductive risk	Additional death rate of females or males because of breeding (e.g. as a result of protecting the offspring against predators). Provide as a vector where the first index indicates the risk for females, the second the risk for males.	From 0 to 1	0.2 for females (Malachowski pers. Comm.)
Mean number of offspring	Mean number of offspring per female	1+	6 (Malachowski et al. 2018)
Sd offspring size	Standard deviation of number of offspring per female (assuming the number of offspring is always 0 or larger)	1+	1 (Malachowski et al. 2018)
Extra Pair Copulation	Probability per offspring to be the result of extra pair copulation	From 0 to 1	0
Maximum age	Organism's maximum age (Default 6)	0+	6 (Malachowski pers. Comm.)
Smin	Minimum survival rate	0+	0.5 (Robinson et al. 2017)
Smax	Maximum survival rate	0+	0.9 (Robinson et al. 2017)
b	Steepness of the survival rate	From -3 to 0	-2 (Robinson et al. 2017)
p	Density at which the survival rate changes most relative.	From 0 to 2	0.5 (Robinson et al. 2017)
Number of generations	Number of generations to simulate	From 2 to 100	20
Target frequency	Target Ancestry	From 0 to 1	0.99

Optimize Supplementation	When set to 0, FALSE or a negative number, it will not be optimized. When negative, the absolute value will be taken as a fixed contribution to each generation (but will not be optimized)	User Preference	TRUE
Optimize Removal	When set to 0, FALSE or a negative number, it will not be optimized. When negative, the absolute value will be taken as a fixed contribution to each generation (but will not be optimized)	User Preference	TRUE
Number of replicates	Number of replicates (bootstraps)	From 0 to 10	100
Verbose	provides verbose output if TRUE.	User preference	User preference
Initial population size	Starting Population size	From 0 to 1000	100
Starting frequency	Initial focal population ancestry frequency in the population	From 0 to 1	Wetland Specific (; Supplemental Table S2.1)
Genetic model	The model can either use a simplified model ("point") of underlying genetics, which speeds up simulation considerably, but underestimates genetic variation. Alternatively, a more detailed genetic model is available, making use of the theory of junctions, this can be accessed using the option "junctions". Default is "simplified".	• point • junctions	junctions
Ancestry put	Average ancestry of individuals being used for supplementation. If the target is high focal ancestry (e.g. aiming for focal ancestry of 1.0), ancestry put should reflect this and be set to 1.0 (which is the default value). When supplementing with non-native individuals, this value can consequently be lowered.	From 0 to 1	1
Ancestry pull	Maximum ancestry of individuals used for pulling	From 0 to 1	1

Sex ratio put	The sex ratio of individuals that are added (if any) to the population. Sex ratio is expressed as males / (males + females), such that 0.5 indicates an even sex ratio, 0.9 indicates a male biased sex ratio and 0.1 indicates a female biased sex ratio.	From 0 to 1	0.5
Sex ratio pull	The sex ratio of individuals that are removed (if any) from the population. The sex ratio is expressed as males / (males + females), such that 0.5 indicates an even sex ratio, 0.9 indicates a male biased sex ratio and 0.1 indicates a female biased sex ratio.	From 0 to 1	0.5
Sex ratio offspring	Sex ratio of newly born offspring. The sex ratio is expressed as males / (males + females), such that 0.5 indicates an even sex ratio, 0.9 indicates a male biased sex ratio and 0.1 indicates a female biased sex ratio	From 0 to 1	0.5
Establishment burn-in	Number of generations since admixture or population establishment.	User preference	30

METHODS

Simulation code for the simRestore package can be accessed directly via the R programming language, and can be incorporated into analysis scripts. It is available as an R package, accessible via CRAN, or via www.github.com/thijsjanzen/simRestore. However, for those unfamiliar with the R programming language the simulation code can also be used via a Shiny Web application (Chang et al., 2022); which can be run either locally, or hosted online. The Shiny Web application presents a user-friendly graphical interface (GUI) in which the user can manipulate conditions through direct value input, buttons, and sliders to adjust chosen parameters (Figure 2.1). Once the parameters are set, the user is then presented with direct graphical feedback indicating the required supplementation or removal efforts, resulting change

in focal ancestry, and population size change. The Shiny Web application is publicly available at https://thijsjanzen.shinyapps.io/simRestoreApp/.

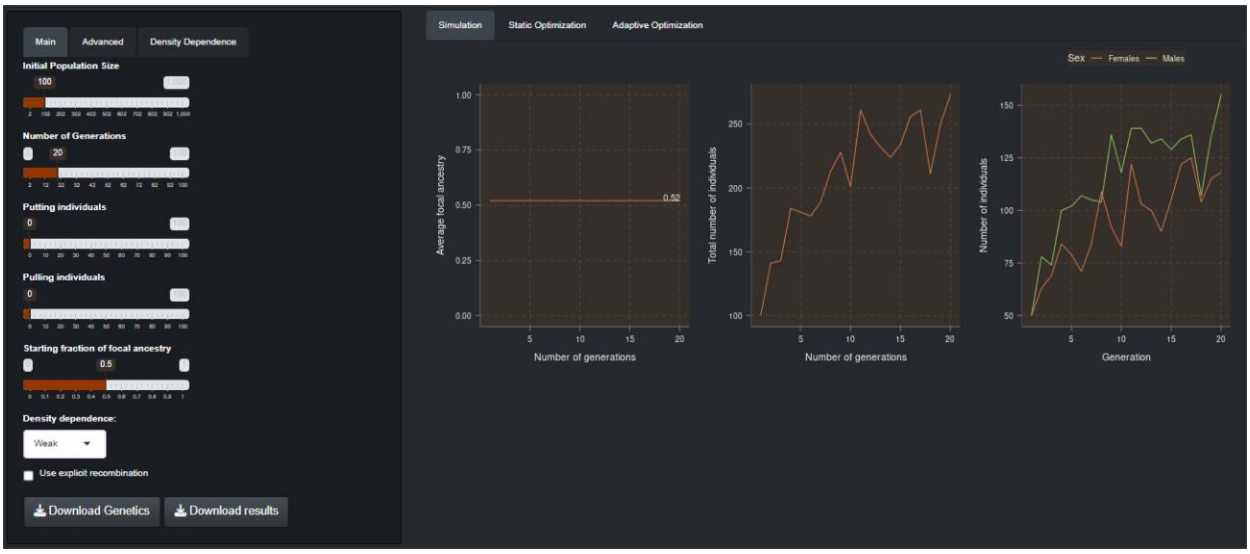

Figure 2.1. Overview of the simRestore shiny app display. (Left) Display of the main variables to modify, and (Right) the three graphs using a simple simulation.

Life-history model

Life history model simulations are implemented using a population with overlapping generations and explicit sexes, where offspring may compete next year for mating opportunities and superfluous males or females may be excluded from mating. I assumed density-dependent population growth, such that over time, an equilibrium density of individuals is reached. The model proceeds through 'seasons' during which the following events occur in sequence: 1) survival, 2) Human intervention, 3) Mating and offspring survival, and 4) Offspring recruitment (Figure 2.2A).

Each season of the life cycle, survival is assumed to be the same and irrespective of sex. However, survival is density-dependent such that at higher densities, individuals have a lower survival rate because of a shortage of resources (Gunnarsson et al., 2013). I model density

dependence of survival following Robinson et al. (2017), who model survival in American Black Duck (*Anas rubripes*) using the following equation:

$$S = S_{max} + \frac{S_{min} - S_{max}}{1 + \left(\frac{D}{p}\right)^{b}}$$

Where *S* is the survival probability (and *max* and *min* indicate the maximum and minimum survival probabilities), *D* is the density (e.g., N / K, where *N* is the number of individuals, and *K* is the carrying capacity), *p* is a variable that indicates the reflection point (e.g., the point at which survival is 50% of S_{max} - S_{min}), and *b* is a variable that indicates the steepness of the curve (Figure 2.2B). In addition to annual survival probability, I also include the capacity for users to include additional mortality for either or both sexes, as breeding animals often have higher rates of mortality (Lima, 2009;Norrdahl & Korpimäki, 1998; Simmons & Kvarnemo, 2006).

Human intervention is modeled as the supplementation or removal of individuals from the population. When supplementing, individuals are by default added in an equal sex ratio (50% males, 50% females), and assuming that the added individuals carry the ancestry of the focal system of interest. When removing, individuals are removed irrespective of sex and can be removed depending on their maximum genetic ancestry. For example, if this variable is chosen to be 0.5, only individuals with less than 50% native species ancestry are removed. By default, the value is chosen to be 1, which means individuals are removed irrespective of their native species ancestry. Additionally, the sex ratio of individuals added or removed can be changed by the user, as can the ancestry of the individuals supplemented.

To simulate mating, I consider two models designed to emulate the diverse range of social and behavioral aspects of reproductions observed across species: 1) “Strict Pair Bonding”,

where females and males pair up in such a way that each female mates with one available male. If there are fewer males than females, some females remain unmated (and vice versa). 2) "Random mating", characterized by mating probabilities being equal among all individuals in a population. Additionally, I introduced a variable accounting for the probability of offspring resulting from EPC (i.e., extra-pair copulation), a behavior primarily observed in birds but also present in other species (Brouwer & Griffith; Soulsbury 2010; Uller & Olsson 2008). Furthermore, I do not model multi-season pair bonding to retain simplicity of the model. Thus, in smaller populations, offspring of a pair might mate with one of their parents in the next season (albeit this chance is relatively low, in the order of 1/N, which is usually less than 1%). Other factors affecting population growth are also considered in the model, such as reproductive success (i.e., the probability of successfully breeding; Table 2.1).

After mating, offspring recruitment is lower than the total number of offspring produced because of a multitude of factors such as predation (natural and human), parasitism, and deteriorating environmental conditions like food abundance (Bortolotti et al., 2011; Gaillard et al., 2008; Hoover & Reetz, 2006; Knights et al., 2012; Rigby, 2008). Thus, I modeled the contribution of variation in recruitment to changes in the population growth as density dependence. For simplicity, I have chosen to make recruitment identical to adult survival and use the same equation (see above equation 1) with identical parameters (Robinson et al., 2017).

Genetic model

The implementation of the genetic model assumes a diploid genome (Zhang et al., 2020), where each chromosome undergoes crossover during meiosis, and the user can specify the total number of chromosomes to be simulated. Upon creation of new offspring, both mother and father provide a single copy of the genome, which are combined to form the diploid offspring

genome. Sex is determined randomly (Trivers & Willard, 1973), independent of the genetic makeup, and ignoring any potential processes driving sex determination or sex skew. Two models are implemented to track ancestry in the genome: 1) an explicit crossover model tracking 'junctions', and 2) a point ancestry model (Figure 2.2C).

In the explicit crossover model, local ancestry is explicitly tracked along the genome and I explicitly model crossovers, resulting recombination events and the effect this has on changes in ancestry. The effects on ancestry are tracked by keeping account of the number of switches in ancestry along the genome, coined 'junctions' by Fisher (1954). Junctions inherit similar to point mutations, provide a computationally efficient method to track changes in ancestry (Baird 1995) and are well studied mathematically (Janzen, Nolte, & Traulsen, 2018, Janzen & Miro Pina, 2022 and references therein). Furthermore, using junctions to track local ancestry (in contrast to, for instance, using high-density SNP maps), allows us to retain high resolution tracking of Linkage Disequilibrium, whilst retaining high computational efficiency (see Janzen & Diaz, 2021 for more expansive scenarios utilizing junctions). Changes in ancestry and the resulting junctions arise from crossover events, and I draw the number of crossovers per chromosome from a Poisson distribution, with the mean rate equal to the size of each chromosome expressed in Morgan, with a default value of 1. The location of each crossover on a chromosome is assumed to be independent of the location of other crossovers on the same chromosome, and crossovers are assumed to be distributed uniformly across the genome. Crossovers are independent between chromosomes. Crossovers on locations with differing ancestry between chromosomes result in recombination and the formation of a junction in the resulting recombined chromosome. Over time, the accumulation of recombinations then creates the ancestry-mosaic observed in admixed individuals.

Alternatively, in the point ancestry model, I only track the average ancestry along the genome and each copy of the genome is represented by a single floating-point number per chromosome, indicating the average proportion of ancestry belonging to either of the two parental species. To create a new copy, the average genome ancestry of both copies per chromosome existing in the parent is used. For example, if the parent has one chromosome copy with 20% focal ancestry, and one chromosome copy with 80% focal ancestry, the resulting single chromosome copy that is being given to the offspring has 50% focal ancestry. Thus, this model assumes that recombination acts as a process that uniformly mixes the two genomes. Although this model assumes an overly simplistic representation of the underlying ancestry dynamics, it provides a computationally efficient method to obtain results that, on average, behave very closely to the junctions method. The reason these two methods are so similar, is that although the accumulation of junctions changes the spatial arrangement of ancestry along a genome, it does not change the average ancestry along the genome, as this is mainly driven by population size effects such as drift. To test whether the genetic ancestry means estimated using the point ancestry and junction genetic model are different, I performed statistical tests using R Statistical Software (v 4.1.3, R Core Team 2023). Shapiro tests were used to evaluate the normal distribution of the data; however, this assumption was not met to perform a parametric test. Therefore, I performed the Wilcoxon rank-sum test to compare my two independent samples.

To initialize simulations, an ancestry mosaic is derived by performing 30 generations of Wright-Fisher admixture (e.g., non-overlapping generations, random mating, constant population size) using the known genetic ancestry of an admixed population as a starting point. I use Wright-Fisher admixture instead of the life-history model previously described to generate the starting mosaic to save computation time, and to constrain the population size such that it

matches the initial population size. For example, if the current frequency is 80/20, then I start the Wright-Fisher admixture with 80% focal ancestry and 20% non-focal ancestry, allowing this "population" to then inbreed over 30 generations to obtain an approximate ancestry mosaic previously shown to generate ancestry blocks of correct size and distribution (Janzen et al., 2018, Janzen & Miró Pina, 2022).

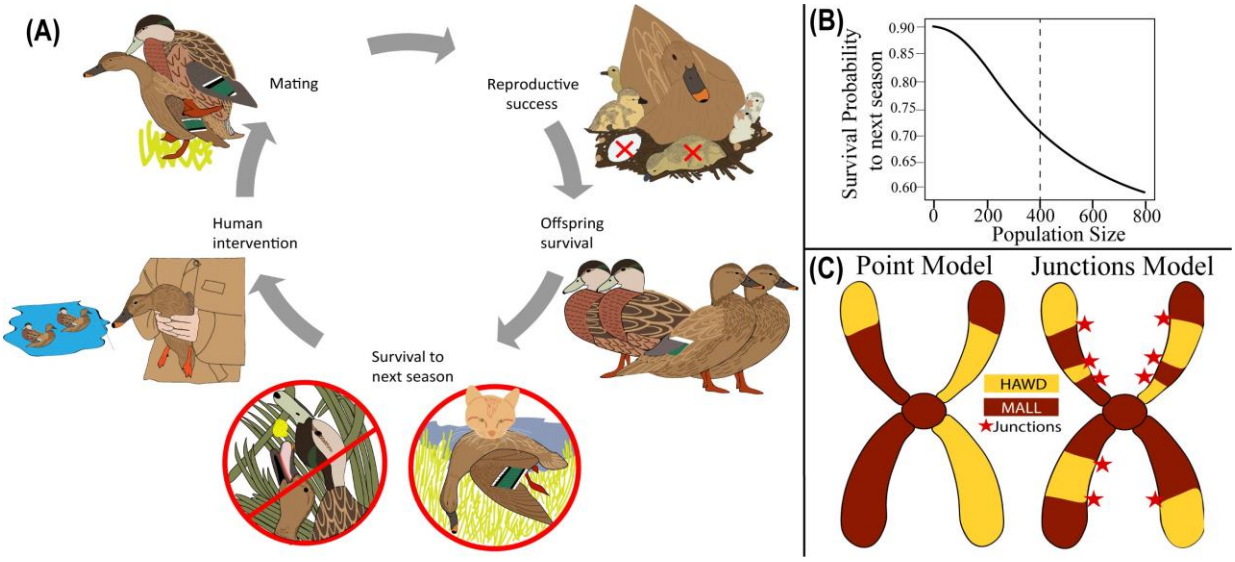

Figure 2.2. Overview of the ecological and genomic information used in the simRestore program. A) Schematic of annual life-history of Hawaiian duck, B) Survival curve based on Robinson et al. (2017) equation, dashed line indicating calculated carrying capacity K = 400, C) Genetic model: Point ancestry model and Junction ancestry model using genetic ancestry of Hawaiian duck (yellow) and mallard (brown).

Static versus Adaptive Simulation Model Optimizations

Apart from providing functionality to forward simulate a chosen set of parameters, the package also provides functionality to optimize parameters to reach a target level of focal ancestry. First, static optimization attempts to find a fixed per-generation amount of input and removal efforts to reach the target level of genetic integrity. Alternatively, an adaptive optimization model provides more specific information by attempting to optimize the

distribution of individuals to be added or removed over a set of generations to reach the target level of genetic integrity. To reduce the degrees of freedom for fitting such a distribution, adaptive optimization does not directly fit the number of individuals per generation (keeping the total constant), but rather fits two parameters of a beta distribution, which is used to determine the number of individuals per generation. Optimization can be performed for supplementation, for removal, or for the joint effort of both supplementation and removal.

Ecological parameter sensitivity analysis

I wanted to understand how the accuracy of the ecological data may impact modeling outputs, and thus, also determine how each ecological parameter influences management strategies. To do so, I tested and ranked how supplementation estimates were affected by successively varying each ecological parameter (Table 2.1). Sensitivity analyses were based on the static optimization of supplementation, assuming a population (*N*) size of 400, a starting genetic ancestry of 20:80 for parental A versus B, with a target ancestry of 99% for focal parental A, and using the junction as the genetic model. Each analysis was run 100 times with optimized supplementation and estimated population sizes following management strategy summarized as boxplots and including the mean value of the percent change in optimized value given a starting value per ecological parameter. In my sensitivity analysis, I first explored the impact of density dependence by varying the steepness / slope of the survival curve (varied as -4, -3.5, …, -0.5; see parameter b in Table 2.1). Similarly, I explored sensitivity of results to the mean number of offspring (i.e., varied as 1, 2, …, 10), maximum individual age (i.e., varied as 1, 2, …, 10), reproductive risk of females and males (i.e., varied as 0, 0.05, …, 0.4), and reproductive success (i.e., varied as 0.0, 0.1, …, 1.0). Specifically, sensitivity ranking of

ecological parameters was based on the magnitude of changes in estimated supplementation number and population size for each assessed parameter value expressed as percentage. The magnitude of change was based on the subtraction of the current estimated number (CE) of individuals/population size and previous estimates (PE), divided by the current value, and all multiplied by 100 % [((CE-PE)/CE)*100%]. For example, to understand differences between the mean number of offspring of two compared to one, the magnitude of change was determined for estimated supplementation and population size as the PE and CE values under the mean number of offspring of one versus two, respectively. In the end, ecological parameters with the largest magnitude of changes between values were considered more sensitive. I plotted and visualized percent change of the absolute values as boxplots using R.

Case study: Simulating Potential Conservation Strategies for Hawaiian Duck

To illustrate the performance of the developed program, I tested whether any set of management efforts could reverse the genetic constitution of the Hawaiian duck x feral mallard hybrids found across Hawaiian Islands. The history and constant monitoring of Hawaiian duck provides high quality ecological and genetic information, including demographic and vital rate differences between native Hawaiian duck on Kauai and hybrids elsewhere, allowing us to optimize management strategies based on wetland-specific conditions (Table 2.1; Supplementary Table S2.1). Note that given that I used the Hawaiian duck as a case example, ecological models were based on their life cycle traits (Table 2.1). All simulations were done with an initial population size (*N*) of 100, and a starting fraction of focal (i.e., Hawaiian duck) ancestry specific to the O‘ahu wetlands Hamukua, Kawainui, and Ki’i (Tables 2.1 & S2.1; Wells, Lavretsky et al. 2019), with a target frequency of 0.99 for the focal ancestry to be achieved over 20 generations,

zero probability of extra-pair copulation, and considering an additional death rate of females during breeding (Table 2.1). Note that I followed Robinson et al. 2017 for density dependence, setting a moderately steep *b* (-2), along with parameter sets as p = 1.0, S_{min}= 0.5, S_{max}= 0.9 (see Figure 2.1B). Both static and adaptive optimization models were run 100 times for each wetland. Moreover, 30 generations was used as this reflects when Hawaiian duck were re-introduced from Kauai into the other Hawaiian Islands; and thus, hybridization with already established feral mallards would have started (Engilis Jr *et al.* 2004).

RESULTS

Genetic and mating models

Whether using the point ancestry or 'junctions' genetic methods, final outputs were statistically identical if inputting values across parameters or optimizing outputs in either the static or adaptive optimization models (Figure 2.3). Importantly, when using the point ancestry or the 'junctions' genetic models, simulations in which no management strategies were undertaken (i.e. no supplementation or removing individuals) resulted in no change in the average local ancestry over time, and the population remained at the starting ancestry proportions (Table S2.1). However, the desired population ancestry goal of 0.99 for the focal population was achieved by supplementing individuals (Table 2.2). Even though the average behavior between the point ancestry and 'junctions' methods was identical, there was higher variance among replicates under the 'junctions' model, indicating that the point model tends to underestimate the observed standing genetic variation (Figure 2.3), as expected.

Next, I found no statistical differences when evaluating between mating models (i.e., struct pair bonding vs. random mating) whether in the absence of a management strategy or

when optimizing the management approach (Figure S2.1). This absence of differentiation between mating models is further reflected in the outcomes of management strategies (i.e., number of supplemented individuals and population size), even when evaluating a spectrum of extra-pair copulation rates ranging from 0 to 1 (Figure S2.2).

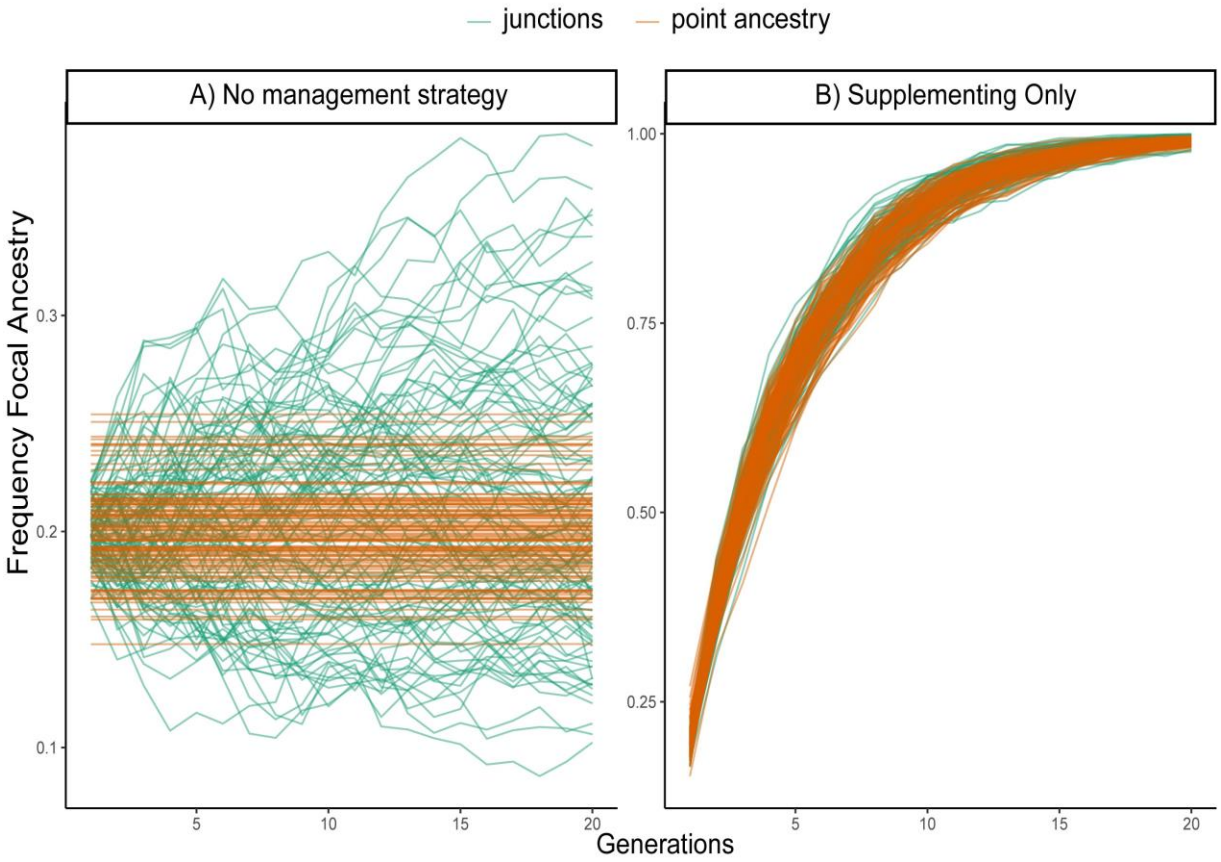

Figure 2.3. Comparing outputs for point ancestry (orange) and the junctions (green) genetic methods based on a (A) no management strategy versus (B) optimization of supplementation only, each run 100 times. As expected, no substantial change between the starting and final ancestries were attained with either genetic method if no management is done, whereas statistically similar (Wilcox-test p-value = 0.95) results of substantial ancestry improvement was recovered with both genetic methods when optimizing supplementation.

Static versus adaptive simulation model optimizations

All analyses were based on the junctions ancestry genetic model. First, optimization of removing only resulted in no change of focal ancestry (e.g., Table 2.2), even when exploring

variations in the genetic ancestry of the removed individuals across different levels of focal species ancestry (Table S2.2). Conversely, optimizing supplementation only reaches the target frequency over the time by consistently adding a greater number of individuals per generation with the static as compared to the adaptive model. However, focal ancestry was always reached with fewer individuals per generation when optimizing both supplementation and removal (e.g., Table 2.2). Applying totals that were recovered in the static simulation into the adaptive simulation model further optimized strategies for each generation (e.g., Table 2.2).

Ecological parameter sensitivity

Sensitivity analyses were based on changes in the optimized number of individuals added to the population, which was estimated by changing each ecological parameter (Figure 2.4) while maintaining the target genetic frequency. Overall, I found that the required individuals to supplement per generation strongly correlated with the obtained population size as permitted by the used parameters that included, (a) lenient parameters allowed for large populations also caused large estimates for the required individuals to supplement, whereas (b) stringent parameter settings causing reduced population sizes and low estimates for the number of required individuals.

At the population level, I found that increasing the strength of density-dependence (e.g. making the slope more negative) resulted in smaller populations and decreasing supplementation requirements (Figure 2.4A). Conversely, with the slope becoming closer to 0.0 and density dependence weakening, I found that population sizes became exponentially larger, and consequently the supplementation requirement per generation increased exponentially as well. Therefore, it seems better to underestimate the strength of density dependence as this results in

overestimating the number to supplement; although, over-supplementation will still result in reaching the target genetic frequency. Next, I find that a shorter life expectancy reduces the required amount to be supplemented, plateauing for any maximum age extending past 6-7 years of age (Figure 2.4C) that is likely because of other mortality effects (density dependence, reproductive risk) acting before the focal individual reaches the maximum life expectancy.

At the mating stage, I find that the mean number of offspring is particularly important, with a 24% error in management strategy if the average number of offspring is inaccurate by even ±1 offspring (Figure 2.4B). As expected, population trends show an exponential increase with each increase in +1 offspring per breeding event. Next, miscalculating additional female or male mortality during reproduction causes deviations in the true supplement number with errors ranging from 3-22% for females and 0-16 % in males (Figure 2.4E, F); .as with density-dependence, overestimating reproductive risk results in reaching the desired focal ancestry albeit over-supplementation. As expected, I find that population trends are inversely associated with increases in mortality of either sex. Finally, both supplemental need and population growth follow an S curve as reproductive success rate increases, with an inflection point for reproductive success of ~50%, and plateauing once reproductive success is ≥80%. Generally, optimized supplementation was underestimated by ~25% for every 0.1% inaccuracy in estimated reproductive success (Figure 2.4D).

The ecological parameters showed different influences on the number of individuals to supplement and the population size in each generation (Figure 2.4G). At the population level, both parameters (i.e. density dependence and maximum age) showed a weak influence in the supplemental of individuals in a population, and therefore the population size. Specifically, the life expectancy parameter had the lowest rank. At the mating stage, the number of offspring and

reproductive success rate showed the strongest influence on the number of individuals needed to add to the population and therefore, the population size. However, the additional mortality rate of males during reproduction had the lowest rank among the parameters that influence the mating stage.

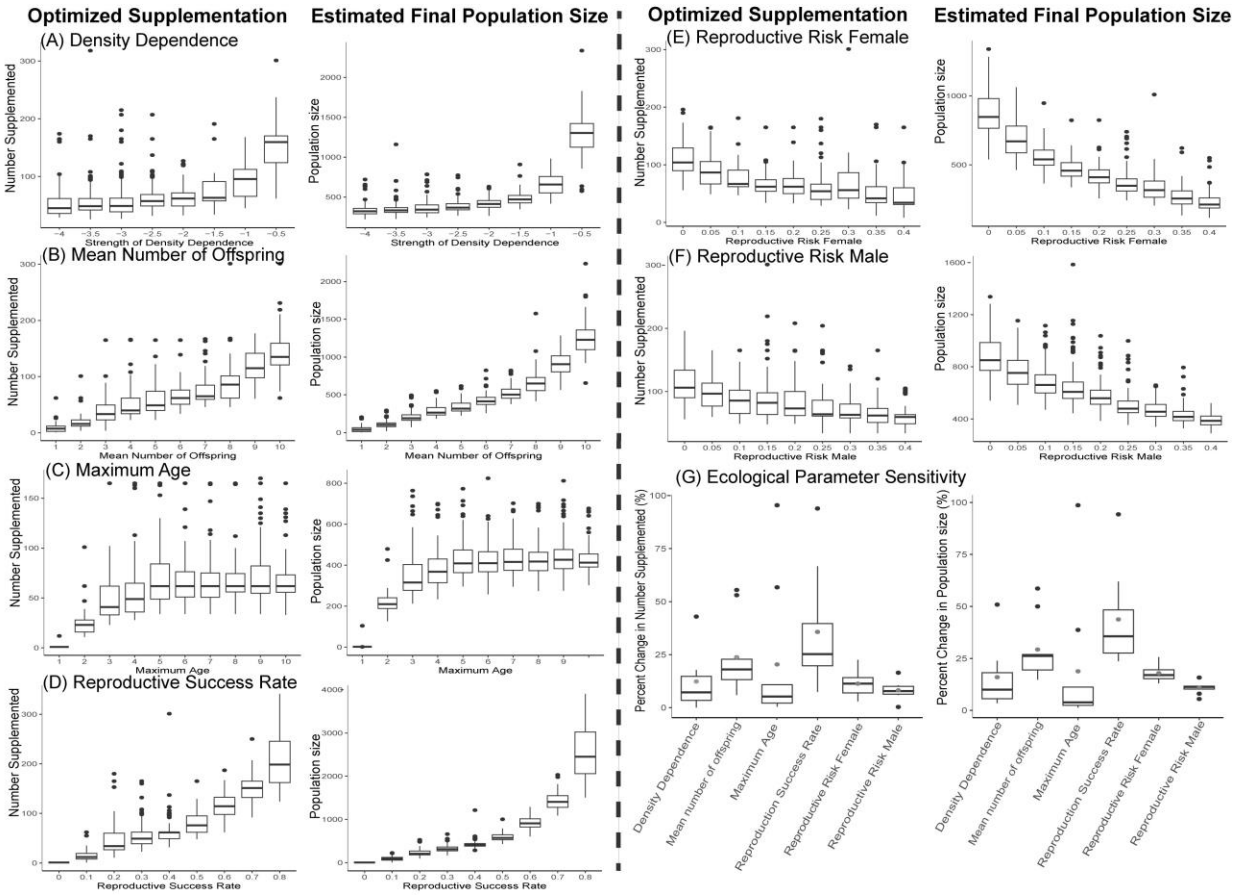

Figure 2.4. (A-F) Changes in optimized supplementation and final population sizes when varying ecological parameters present in the 'simRestore' R package. (G) Magnitude of change expressed as percentage when varying A-F ecological variables in estimating supplementation and population size and depicted as box plots with mean values denoted (grey circle).

By default, the sex ratio of offspring is equal (e.g. offspring has a 50% probability to be of the male sex). Deviations from this even sex ratio significantly influenced outcomes if

populations are strongly sex biased (e.g. 0.9 indicates a male biased sex ratio and 0.1 indicates a female biased sex ratio; Figure 2.5 B,D). Similarly, the number of individuals required to supplement also deviated if the sex ratio of the supplementation population was not even (Figure 2.5 A,C). Male biased supplementation resulted in overall population size declines, whereas female biased input resulted in population growth. Furthermore, female biased supplementation caused an increase in the number of individuals required to be put to reach the target genetic frequency, whereas male biased supplementation reduced this. Thus, the sex composition of the supplementing population can be skewed towards male or female if managers are attempting to simply change the genetic ancestry or also increase their population size, respectively.

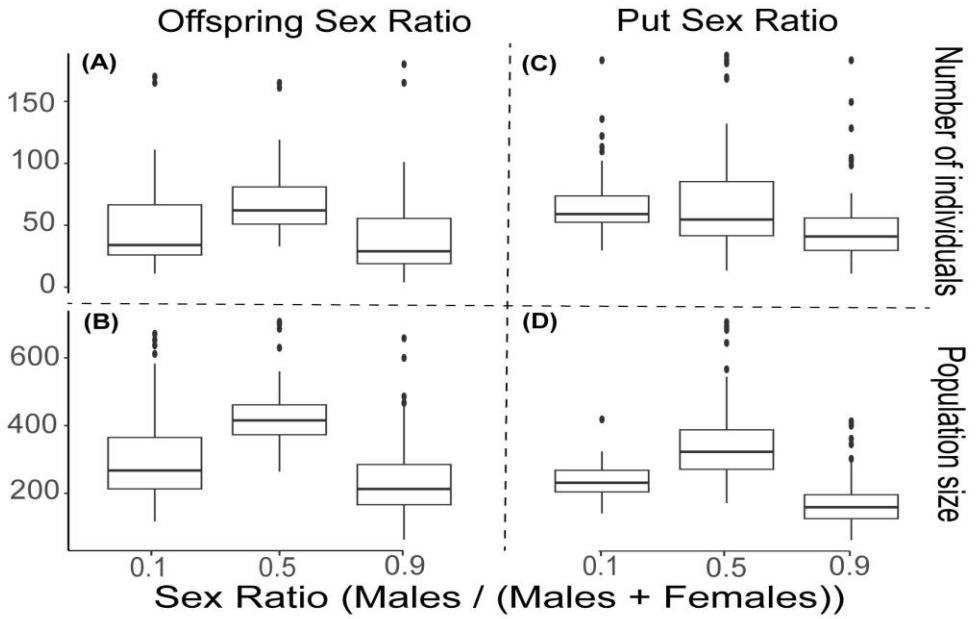

Figure 2.5. Changes in optimized supplementation (top row) and final population sizes (bottom row) when varying sex ratios of offspring (first column) or individuals added (second column) in the 'simRestore' R package. Sex ratio is expressed as males / (males + females), such that 0.5 indicates an even sex ratio, 0.9 indicates a male biased sex ratio and 0.1 indicates a female biased sex ratio.

Simulating potential Hawaiian duck conservation strategies

I simulated optimized management strategies to reverse the population's genetic ancestry for several wetlands on O'ahu, Hawaii with varying starting average Hawaiian duck ancestry (Table 2.2). First, however, all simulations clearly show that the removal of individuals does not substantially change the focal ancestry (Table 2.2). More importantly, whereas management strategies employing supplementation only eventually reach ancestry goals, these require substantially more individuals than what would be required if strategies included a mix of supplementation and removal efforts (Table 2.2). Finally, by varying the management strategy generationally, the adaptive model generally decreases the number of generations to reach ancestry goal as compared to the static model (Table 2.2). For example, static simulations suggest that the ancestry can be moved from 74% to 99% Hawaiian duck on O'ahu's Ki'i wetland complex by adding 60 Hawaiian ducks and removing 3 hybrids per generation for 20 generations. Conversely, applying the adaptive simulation model reaches 99% Hawaiian duck ancestry in five generations by varying the supplementation and removal strategies across generations (Table 2.2).

Table 2.2. Optimized management strategies for three wetlands on O‘ahu of varying starting average Hawaiian duck ancestry (Table 2.1). Static and adaptive optimizations were done under management strategies including supplementation only, removing only, or a combination of supplementation and removal. Success of each strategy was determined by the predicted final ancestry, with the objective set to reach 99% Hawaiian duck ancestry. Note that while static optimization simply allocates the same number of individuals across set generations (gen), adaptive optimization varies the strategy across generations.

Model	Management Strategy	Wetland (Current Avg. Hawaiian duck Ancestry of Population)		
		Hamukua (28%)	**Kawainui (48%)**	**Ki'i (74%)**
Static Optimization	Supplement Only	Supplement 93 Hawaiian ducks/gen x 20 gens = **1,860 Hawaiian ducks** Final Ancestry = 0.999 = **Achieved**	Supplement 84 Hawaiian ducks /gen x 20 gens = **1,680 Hawaiian ducks** Final Ancestry = 0.999 = **Achieved**	Supplement 75 Hawaiian ducks /gen x 20 gens = **1,500 Hawaiian ducks** Final Ancestry = 0.999 = **Achieved**
	Removing Only	Remove 13 Hybrids/gen x 20 gens = **260 Hybrids** Final Ancestry = 0.282 = **Failed**	Remove 19 Hybrids /gen x 20 gens = **380 Hybrids** Final Ancestry = 0.506 = **Failed**	Remove 21 Hybrids /gen x 20 gens = **420 Hybrids** Final Ancestry = 0.791 = **Failed**
	Supplement + Removing	Supplement 64 Hawaiian ducks /gen x 20 gens = **1,280 Hawaiian ducks** + Remove 23 Hawaiian ducks /gen x 20 gens = **460 Hybrids** Final Ancestry = 0.999 = **Achieved**	Supplement 76 Hawaiian ducks /gen x 20 gens = **1,520 Hawaiian ducks** + Remove 5 Hawaiian ducks /gen x 20 gens = **100 Hybrids** Final Ancestry = 0.999 = **Achieved**	Supplement 60 Hawaiian ducks /gen x 20 gens = **1,200 Hawaiian ducks** + Remove 3 Hawaiian ducks /gen x 20 gens = **60 Hybrids** Final Ancestry = 0.999 = **Achieved**
Adaptive Optimization	Supplement Only	Supplement 748 total Hawaiian ducks Vary strategy for 19 generations Final Ancestry = 0.986 = **Nearing**	Supplement 647 total Hawaiian ducks Vary strategy for 19 generations Final Ancestry = 0.986 = **Nearing**	Supplement 517 total Hawaiian ducks Vary strategy for 19 generations Final Ancestry = 0.99 = **Achieved**
	Removing Only	Remove 104 total Hybrids Vary strategy for 9 generations Final Ancestry = 0.306 = **Failed**	Remove 20 total Hybrids Vary strategy for 1 generations Final Ancestry = 0.503 = **Failed**	Remove 10 total Hybrids Vary strategy for 2 generations Final Ancestry = 0.807 = **Failed**
	Supplement + Removing	Supplement 110 total Hawaiian ducks + Remove 50 total Hybrids Vary strategy for 12 generations Final Ancestry = 1.000 = **Achieved**	Supplement 120 total Hawaiian ducks + Remove 20 total Hybrids Vary strategy for 12 generations Final Ancestry = 0.999 = **Achieved**	Supplement 97 total Hawaiian ducks + Remove 30 total Hybrids Vary strategy for 5 generations Final Ancestry = 1.000 = **Achieved**

DISCUSSION

My software package simRestore uses a forward simulator platform that optimizes available management strategies to meet set conservation goal(s). I provide functionality to the program through a user-friendly GUI interface where anyone can change ecological parameters and genetic ancestry information to forward simulate actionable management strategies. Importantly, simRestore was developed to permit for management strategy optimization at any geographical (i.e., wetland, Island) and temporal (i.e., number of generations) scale, as well as under any project limitations (i.e., number of individuals available for input, time for project completions). Finally, assessing offspring ancestry allows for a direct test of the implemented strategy; and thus, permitting researchers to determine ecological or other factors that were not accurately accounted for in situations where simulated values are unaligned to empirical ones. Assessing simulation accuracy can be done annually by comparing genetic ancestry of offspring in the simulation with genetic ancestry assessed empirically. If genetic integrity is reversed but set ancestry goals are not attained following the required number of generations of stocking, then the newly established genetic information can be fed back into models to help guide the following year's strategies. In the end, the simRestore program provides a means for the adaptive management planning for species' conservation.

Model considerations

Although the point and junctions genetic models resulted in identical average behavior in resulting ancestry, there was a higher degree of variation when using the junctions model (Figure 2.3). I conclude that whereas the point model is a strong oversimplification of the underlying genetics, its fast computation and high similarity in outcome provides a great benefit over the

more complex and demanding junctions model. I suggest the use of the point model in management strategy optimization followed by an assessment in outcome variability and robustness with the junctions model. More importantly, for organisms for which more detailed genomic information remains lacking (i.e. chromosome size, recombination rate, etc), and thus the use of the junctions genetic model is perhaps inappropriate, the point genetic model will still provide robust inferences. Similarly, varying both the mating models (strict and pair-bonded) and the EPC variable produced comparable changes in ancestry and management strategies, as they are not inherently tied to a particular ancestry type (e.g., males of native ancestry show a greater likelihood of involvement in EPC). However, I strongly recommend setting up the mating system of the studied organism to the best of my knowledge to ensure inference reliability.

Next, assessing the utility of the three management strategies, I conclude that removing only is similar to a no management strategy as the local ancestry is unaffected (Table 2.2; Table S2.2). This is not unsurprising, as removing individuals of a hybrid swarm still results in remaining breeding individuals being hybrid; confirming that a parental gene pool is required to reverse the genetic ancestry of a hybrid pool (Wells et al., 2019). Whereas supplementing only can reverse the genetic signature towards the focal species, the total number required to be supplemented is consistently higher and requiring more generations than when combining supplementation and removal efforts (Table 2.2; Table S2.2). Thus, the optimum management strategy is combining supplementation and removal when possible.

In addition to starting genetic ancestry, researchers will also need to know a suite of ecological parameters for their population(s) of interest. Life-history traits can influence how much maternal vs. paternal genetic variation contributes to each generation (Table 2.1). Life-history traits therefore need to be correctly incorporated to ensure accurate model predictions. I

find that each of the ecological parameters within my model affects simulations differently, and with some being largely insensitive, while others are highly sensitive and require accurate estimates (Figures 2.4 & 2.5). Among parameters, researchers need to take particular care to have accurate estimates of the number of offspring (i.e., annual fecundity) and reproductive success rate in particular, as these have a disproportionate effect on population size, which in turn dictates management strategies. Moreover, researchers working in a female biased population and/or r-selected species will require substantial increases in supplemental efforts (Figure 2.5). More generally, however, I find that miscalculating most ecological parameters may not impact management outcomes, since doing so still typically results in the achievement of management goals, albeit on slightly different time-frames and/or supplementation efforts. Specifically, unless known, I recommend that it is better to underestimate density dependency, overestimate reproductive risk, and assume a 50:50 number of offspring as all of these will result in optimum management strategies that can still achieve ancestry goals (Figure 2.4 & 2.5); though at the expense of additional resources that would otherwise not be required.

Conservation implications

The incorporation of genetic data has become a fundamental source of information for species conservation (Walters & Schwartz, 2020). Genomic data has been used to shed light into species' effective population size, inbreeding demographic history, and population structure that not only aid in management efforts (Hohenlohe et al., 2021), but are critical in efforts of biodiversity monitoring, resolving taxonomic uncertainty, wildlife forensics, and designation of conservation units (Funk et al., 2012; Hohenlohe et al., 2021). However, there has been a severe lag and lack in the implementation of genomic data into management decisions (Walters & Schwartz, 2020); in part because of the lack of user-friendly methods (Hohenlohe et al., 2021).

From the perspective of wildlife management, understanding the population genomics of wild populations can aid multiple traditional wildlife activities such as translocations, reintroductions, population augmentation, and the identification of units of conservation concern (Funk et al., 2012; Hohenlohe et al., 2021; Tallmon et al., 2004; Walters & Schwartz, 2020; Whiteley et al., 2015). Among conservation efforts, understanding rates of hybridization and its implication to wildlife populations is now at the forefront of conservation science (Cooper & Shaffer, 2021; Hohenlohe et al., 2021; Searcy et al., 2016). Towards these efforts, my developed simRestore program provides a tool where managers can couple genetic and ecological data to optimize management strategies that directly feed into decision making when attempting to resolve hybridization issues. Once again, the ability to feed empirical data (i.e., genetic assignment probabilities of a population) back into models provide a powerful tool for management to be adaptive and specific to the species' needs.

I apply developed methods to understand the potential in artificially reversing genetic ancestry through directed management efforts for the endangered Hawaiian duck. Given the proximate threat to the Hawaiian duck is genetic extinction through ongoing anthropogenic hybridization with feral mallards (USFWS, 2012; Wells et al., 2019), I demonstrate that the reversal to a genetic native status of these hybrid swarms is theoretically possible (Table 2.2). Importantly, both molecular data (Lavretsky et al., 2019; Wells et al., 2019) and telemetry movement data (Malachowski, 2013; Malachowski et al., 2019, 2020; Malachowski & Dugger, 2018) suggest that movement is not only limited between Islands, but also among wetlands within Islands. Consequently, each wetland on Islands can be effectively considered a closed system, providing an important means to closely monitor how optimized management strategies impact the genetic integrity of that wetland's population (Table 2.2). I conclude that the

developed models can help guide future Hawaiian duck conservation efforts, with work the implementation of optimized strategies in a pilot program to determine real-world feasibility of the simulations.

CONCLUSSIONS

Conceptually, hybrid individuals are simply conduits of genes stemming from divergent lineages (Allendorf et al., 2001); and as a result, hybrids still possess the genetic diversity of the species of interest. Though hybrids pose a conservation concern because of the maladaptive potential when divergent genomes are admixed (i.e., outbreeding; Templeton, 1986), the genetic diversity of the focal species remains present. Given that species of conservation concern often suffer in population size and standing genetic diversity, continued losses in both are often detrimental. Here, I demonstrate that directed manipulation of hybrid populations can potentially reconstitute a hybrid population towards the focal population without additional loss of individuals and the genetic diversity they carry. However, what wildlife biologists are able to do to mitigate or reverse the continued loss of individuals to hybridization can be further limited by human dimensions (i.e., social rejection of particular management strategies; (Decker et al., 2012). Thus, the functionality of the simRestore program provides biologists and managers attempting to mitigate or reverse such trends, a decision-making tool for management optimization specific to the species or population of concern in an adaptive framework. By doing so, biologists not only have the means to evaluate possible conservation scenarios but critical information to explain and with which to engage their constituents.

ACKNOWLEDGMENTS

I are thankful to Afsheen A. Siddiqi, Annie Marshall, Kelly Goodale, Kim Uyehara, Andrew Engilis Jr., Caitlin P. Wells, Christopher P. Malachowski, Bruce D. Dugger, and John M. Eadie for data and suggestions in building Hawaiian duck models. This research was funded by Pacific Island U.S. Fish and Wildlife Services.

SIMULATION CODE AND DATA ACCESIBILITY

The simulation code described in this manuscript is available as an R package, accessible via CRAN, via github (www.github.com/thijsjanzen/simRestore), or via the Shiny app on https://thijsjanzen.shinyapp.sio/simRestoreApp. All simulations performed in this manuscript were performed using genetic ancestry data presented in Table S2.1.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013a). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013b). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Abbott, R. J., Hegarty, M. J., Hiscock, S. J., & Brennan, A. C. (2010). Homoploid hybrid speciation in action. *TAXON*, *59*(5), 1375–1386. https://doi.org/10.1002/tax.595005

Alerstam, T., & Högstedt, G. (1982). Bird Migration and Reproduction in Relation to Habitats for Survival and Breeding. *Ornis Scandinavica (Scandinavian Journal of Ornithology)*, *13*(1), 25–37. https://doi.org/10.2307/3675970

Aliabadian, M., & Nijman, V. (2007). Avian hybrids: Incidence and geographic distribution of hybridisation in birds. *Contributions to Zoology*, *76*(1), 59–61. https://doi.org/10.1163/18759866-07601006

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*(3), 1217–1229.

Arnold, T. W., & Clark, R. G. (1996). Survival and Philopatry of Female Dabbling Ducks in Southcentral Saskatchewan. *The Journal of Wildlife Management*, *60*(3), 560–568. https://doi.org/10.2307/3802073

Athens, J. S., Toggle, H. D., Ward, J. V., & Welch, D. J. (2002). Avifaunal extinctions, vegetation change, and Polynesian impacts in prehistoric Hawai'i. *Archaeology in Oceania*, *37*(2), 57–78. https://doi.org/10.1002/j.1834-4453.2002.tb00507.x

Barth, J. M. I., Gubili, C., Matschiner, M., Tørresen, O. K., Watanabe, S., Egger, B., Han, Y.-S., Feunteun, E., Sommaruga, R., Jehle, R., & Schabetsberger, R. (2020). Stable species boundaries despite ten million years of hybridization in tropical eels. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15099-x

Barton, N. H. (2013a). Does hybridization influence speciation? *Journal of Evolutionary Biology*, *26*(2), 267–269. https://doi.org/10.1111/jeb.12015

Barton, N. H. (2013b). Does hybridization influence speciation? *Journal of Evolutionary Biology*, *26*(2), 267–269. https://doi.org/10.1111/jeb.12015

Beadell, J. S., Ishtiaq, F., Covas, R., Melo, M., Warren, B. H., Atkinson, C. T., Bensch, S., Graves, G. R., Jhala, Y. V., Peirce, M. A., Rahmani, A. R., Fonseca, D. M., & Fleischer, R. C. (2006). Global phylogeographic limits of Hawaii's avian malaria. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1604), 2935–2944. https://doi.org/10.1098/rspb.2006.3671

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bortolotti, L. E., Harriman, V. B., Clark, R. G., & Dawson, R. D. (2011). Can changes in provisioning by parent birds account for seasonally declining patterns of offspring recruitment? *Canadian Journal of Zoology*, *89*(10), 921–928. https://doi.org/10.1139/z11-068

Boyer, A. G. (2008). Extinction patterns in the avifauna of the Hawaiian islands. *Diversity and Distributions*, *14*(3), 509–517. https://doi.org/10.1111/j.1472-4642.2007.00459.x

Brennan, A., Woodward, G., Seehausen, O., Muñoz-Fuentes, V., Moritz, C., Guelmami, A., Abbott, R., & Edelaar, P. (2014). Hybridization due to changing species distributions: Adding problems or solutions to conservation of biodiversity during global change? *Evolutionary Ecology Research*, *16*.

Brockmeier, E. K., Scott, P. D., Denslow, N. D., & Leusch, F. D. L. (2016). Transcriptomic and physiological changes in Eastern Mosquitofish (*Gambusia holbrooki*) after exposure to progestins and anti-progestagens. *Aquatic Toxicology*, *179*, 8–17. https://doi.org/10.1016/j.aquatox.2016.08.002

Brown, J. I., Harrigan, R. J., & Lavretsky, P. (2022). Evolutionary and ecological drivers of local adaptation and speciation in a North American avian species complex. *Molecular Ecology*, *31*(9), 2578–2593. https://doi.org/10.1111/mec.16423

Browne, R. A., Griffin, C. R., Chang, P. R., Hubley, M., & Martin, A. E. (1993). Genetic Divergence Among Populations Of The Hawaiian Duck, Laysan Duck, And Mallard. *The Auk*, *110*(1), 49–56. https://doi.org/10.1093/auk/110.1.49

Cantabella, E., Camilleri, V., Cavalie, I., Dubourg, N., Gagnaire, B., Charlier, T. D., Adam-Guillermin, C., Cousin, X., & Armant, O. (2022). Revealing the Increased Stress Response Behavior through Transcriptomic Analysis of Adult Zebrafish Brain after Chronic Low to Moderate Dose Rates of Ionizing Radiation. *Cancers*, *14*(15), Article 15. https://doi.org/10.3390/cancers14153793

Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2020). *Genome-wide local ancestries discriminate homoploid hybrid speciation from secondary introgression in the red wolf (Canidae: Canis rufus)* (p. 2020.04.05.026716). bioRxiv. https://doi.org/10.1101/2020.04.05.026716

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2022). Shiny: Web application framework for R. R package version 1.7.3. *URL Https://CRAN. R-Project. Org/Package= Shiny*.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*(1), 128. https://doi.org/10.1186/1471-2105-14-128

Cooper, R. D., & Shaffer, H. B. (2021). Allele-specific expression and gene regulation help explain transgressive thermal tolerance in non-native hybrids of the endangered California tiger salamander (Ambystoma californiense). *Molecular Ecology*, *30*(4), 987–1004. https://doi.org/10.1111/mec.15779

Crispo, E., Moore, J.-S., Lee-Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals. *BioEssays*, *33*(7), 508–518. https://doi.org/10.1002/bies.201000154

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Decker, edited by D. J., Riley, S. J., & Siemer, and W. F. (2012). *Human Dimensions of Wildlife Management*. https://www.press.jhu.edu/books/title/10090/human-dimensions-wildlife-management

Drilling, N., Titman, R. D., & McKinney, F. (2020). Mallard (Anas platyrhynchos), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.mallar3.01species_shared.bow.project_name

Durand, E. Y., Patterson, N., Reich, D., & Montgomery, S. (2011). *Testing for Ancient Admixture between Closely Related Populations*. *28*(8), 2239–2252. https://doi.org/10.1093/molbev/msr048

Elgvin, T. O., Trier, C. N., Tørresen, O. K., Hagen, I. J., Lien, S., Nederbragt, A. J., Ravinet, M., Jensen, H., & Sætre, G.-P. (2017). The genomic mosaicism of hybrid speciation. *Science Advances*, *3*(6), e1602996. https://doi.org/10.1126/sciadv.1602996

Engilis, A., & Pratt, T. (1993). Status and Population Trends of Hawaii's Native Waterbirds, 1977-1987. *The Wilson Bulletin*, *105*, 142–158.

Engilis Jr., A., Uyehara, K. J., & Giffin, J. G. (2020a). Hawaiian Duck (Anas wyvilliana), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.hawduc.01species_shared.bow.project_name

Engilis Jr., A., Uyehara, K. J., & Giffin, J. G. (2020b). Hawaiian Duck (Anas wyvilliana), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.hawduc.01

Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, *11*(7), 1035–1052. https://doi.org/10.1111/eva.12569

Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610–2618. https://doi.org/10.1111/mec.13139

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gaillard, J. M., Coulson, T., & Festa-Bianchet, M. (2008). Recruitment. In S. E. Jørgensen & B. D. Fath (Eds.), *Encyclopedia of Ecology* (pp. 2982–2986). Academic Press. https://doi.org/10.1016/B978-008045405-4.00655-8

Gill, F. B. (2014). *Species taxonomy of birds: Which null hypothesis?Taxonomía de especies de aves: ¿Cuál es la hipótesis nula?Species taxonomy of birds | Ornithology | Oxford Academic*. *131*(2), 150–161.

Grant, P. R., & Grant, B. R. (1992a). Hybridization of bird species. *Science (New York, N.Y.)*, *256*(5054), 193–197. https://doi.org/10.1126/science.256.5054.193

Grant, P. R., & Grant, B. R. (1992b). Hybridization of Bird Species. *Science*, *256*(5054), 193–197. https://doi.org/10.1126/science.256.5054.193

Grapputo, A., Thrimawithana, A. H., Steinwender, B., & Newcomb, R. D. (2018). Differential gene expression in the evolution of sex pheromone communication in New Zealand's endemic leafroller moths of the genera Ctenopseustis and Planotortrix. *BMC Genomics*, *19*(1), 94. https://doi.org/10.1186/s12864-018-4451-1

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science (New York, N.Y.)*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Gunnarsson, G., Elmberg, J., Pöysä, H., Nummi, P., Sjöberg, K., Dessborn, L., & Arzel, C. (2013). Density dependence in ducks: A review of the evidence. *European Journal of Wildlife Research*, *59*(3), 305–321. https://doi.org/10.1007/s10344-013-0716-9

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hedrick, P. W., & Fredrickson, R. (2010). Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics*, *11*(2), 615–626. https://doi.org/10.1007/s10592-009-9999-5

Hennessy, S. M., Wisinski, C. L., Ronan, N. A., Gregory, C. J., Swaisgood, R. R., & Nordstrom, L. A. (2022). Release strategies and ecological factors influence mitigation translocation outcomes for burrowing owls: A comparative evaluation. *Animal Conservation*, *25*(5), 614–626. https://doi.org/10.1111/acv.12767

Hirashiki, C., Kareiva, P., & Marvier, M. (2021). Concern over hybridization risks should not preclude conservation interventions. *Conservation Science and Practice*, *3*(4), e424. https://doi.org/10.1111/csp2.424

Hoban, S., Gaggiotti, O., Consortium, C., & Bertorelle, G. (2013). Sample Planning Optimization Tool for conservation and population Genetics (SPOTG): A software for choosing the appropriate number of markers and samples. *Methods in Ecology and Evolution*, *4*(3), 299–303. https://doi.org/10.1111/2041-210x.12025

Hoff, K. J., & Stanke, M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, *65*(1), e57. https://doi.org/10.1002/cpbi.57

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62–82. https://doi.org/10.1111/mec.15720

Hoover, J. P., & Reetz, M. J. (2006). Brood Parasitism Increases Provisioning Rate, and Reduces Offspring Recruitment and Adult Return Rates, in a Cowbird Host. *Oecologia*, *149*(1), 165–173.

Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, *27*(23), 4839–4855. https://doi.org/10.1111/mec.14858

Janzen, T., Nolte, A. W., & Traulsen, A. (2018). The breakdown of genomic ancestry blocks in hybrid lineages given a finite number of recombination sites. *Evolution; International Journal of Organic Evolution*, *72*(4), 735–750. https://doi.org/10.1111/evo.13436

Jiggins, C. D., Salazar, C., Linares, M., & Mavarez, J. (2008). Hybrid trait speciation and Heliconius butterflies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1506), 3047–3054. https://doi.org/10.1098/rstb.2008.0065

Johnsgard, P. A. (1960). Hybridization in the Anatidae and Its Taxonomic Implications. *The Condor*, *62*(1), 25–33. https://doi.org/10.2307/1365656

Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research*, *30*(1), 42–46. https://doi.org/10.1093/nar/30.1.42

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., Marzluff, J. M., Fleischer, R. C., Johnsen, A., & Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, *9*(1), 906. https://doi.org/10.1038/s41467-018-03294-w

Knights, A. M., Firth, L. B., & Walters, K. (2012). Interactions between Multiple Recruitment Drivers: Post-Settlement Predation Mortality and Flow-Mediated Recruitment. *PLOS ONE*, *7*(4), e35096. https://doi.org/10.1371/journal.pone.0035096

Kour, A., Deb, S. M., Nayee, N., Niranjan, S. K., Raina, V. S., Mukherjee, A., Gupta, I. D., & Patil, C. S. (2023). Novel insights into genome-wide associations in Bos indicus reveal genetic linkages between fertility and growth. *Animal Biotechnology*. https://www.tandfonline.com/doi/abs/10.1080/10495398.2021.1932520

Lamichhaney, S., Han, F., Webster, M. T., Andersson, L., Grant, B. R., & Grant, P. R. (2018). Rapid hybrid speciation in Darwin's finches. *Science (New York, N.Y.)*, *359*(6372), 224–228. https://doi.org/10.1126/science.aao4593

Lavretsky, P. (2021). Population Genomics Provides Key Insights into Admixture, Speciation, and Evolution of Closely Related Ducks of the Mallard Complex. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population Genomics: Wildlife* (pp. 295–330). Springer International Publishing. https://doi.org/10.1007/13836_2020_76

Lavretsky, P., Engilis Jr, A., Eadie, J. M., & Peters, J. L. (2015). Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. *Journal of Evolutionary Biology*, *28*(5), 1005–1015. https://doi.org/10.1111/jeb.12637

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. (2023a). *Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos)*. *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. E. (2023b). Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos). *G3 Genes|Genomes|Genetics*, *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, *9*(6), 3470–3490. https://doi.org/10.1002/ece3.4981

Lavretsky, P., McCracken, K. G., & Peters, J. L. (2014). Phylogenetics of a recent radiation in the mallards and allies (Aves: Anas): Inferences from a genomic transect and the multispecies coalescent. *Molecular Phylogenetics and Evolution*, *70*, 402–411. https://doi.org/10.1016/j.ympev.2013.08.008

Lavretsky, P., McInerney, N. R., Mohl, J. E., Brown, J. I., James, H. F., McCracken, K. G., & Fleischer, R. C. (2020). Assessing changes in genomic divergence following a century of human-mediated secondary contact among wild and captive-bred ducks. *Molecular Ecology*, *29*(3), 578–595. https://doi.org/10.1111/mec.15343

Lavretsky, P., Peters, J. L., Winker, K., Bahn, V., Kulikova, I., Zhuravlev, Y. N., Wilson, R. E., Barger, C., Gurney, K., & McCracken, K. G. (2016). Becoming pure: Identifying generational classes of admixed individuals within lesser and greater scaup populations. *Molecular Ecology*, *25*(3), 661–674. https://doi.org/10.1111/mec.13487

Lee, S. Y., Sohn, K.-A., & Kim, J. H. (2012). MicroRNA-centric measurement improves functional enrichment analysis of co-expressed and differentially expressed microRNA clusters. *BMC Genomics*, *13*(7), S17. https://doi.org/10.1186/1471-2164-13-S7-S17

Leitwein, M., Gagnaire, P.-A., Desmarais, E., Berrebi, P., & Guinand, B. (2018). Genomic consequences of a recent three-way admixture in supplemented wild brown trout populations revealed by local ancestry tracts. *Molecular Ecology*, *27*(17), 3466–3483. https://doi.org/10.1111/mec.14816

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), Article 7357. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Livezey, B. C. (1993). Comparative morphometrics of Anas ducks, with particular reference to the Hawaiian Duck Anas wyvilliana, Laysan Duck A. laysanensis, and Eaton's Pintail A. eatoni. *Wildfowl*, *44*(44), Article 44.

López-Caamal, A., & Tovar-Sánchez, E. (2014). Genetic, morphological, and chemical patterns of plant hybridization. *Revista Chilena de Historia Natural*, *87*(1), 16. https://doi.org/10.1186/s40693-014-0016-0

Malachowski, C. P. (2013). *Hawaiian duck (Anas wyvilliana) behavior and response to wetland habitat management at Hanalei National Wildlife Refuge on Kaua'i*. Oregon State University.

Malachowski, C. P., & Dugger, B. D. (2018). Hawaiian Duck Behavioral Patterns in Seasonal Wetlands and Cultivated Taro. *The Journal of Wildlife Management*, *82*(4), 840–849.

Malachowski, C. P., Dugger, B. D., & Heard, D. J. (2020). Extrusion of Intracoelomic Radiotransmitters by Hawaiian Ducks. *Wildlife Society Bulletin*, *44*(4), 741–748. https://doi.org/10.1002/wsb.1134

Malachowski, C. P., Dugger, B. D., & Uyehara, K. J. (2019). Seasonality of Life History Events and Behavior Patterns in the Island Endemic Hawaiian Duck (Anas wyvilliana). *Waterbirds*, *42*(1), 78–89. https://doi.org/10.1675/063.042.0109

Malachowski, C. P., Dugger, B. D., Uyehara, K. J., & Reynolds, M. H. (2022). Avian botulism is a primary, year-round threat to adult survival in the endangered Hawaiian Duck on Kaua‘i, Hawai‘i, USA. *Ornithological Applications*, *124*(2), duac007. https://doi.org/10.1093/ornithapp/duac007

Mallet, J. (2005a). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2005b). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007a). Hybrid speciation. *Nature*, *446*(7133), Article 7133. https://doi.org/10.1038/nature05706

Mallet, J. (2007b). Hybrid speciation. *Nature*, *446*(7133), Article 7133. https://doi.org/10.1038/nature05706

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature*, *441*(7095), 868–871. https://doi.org/10.1038/nature04738

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

Meguid, M. M., Fetissov, S. O., Blaha, V., & Yang, Z. J. (2000). Dopamine and serotonin VMN release is related to feeding status in obese and lean Zucker rats. *Neuroreport*, *11*(10), 2069–2072. https://doi.org/10.1097/00001756-200007140-00002

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/ncomms14363

Miller, J. M., Poissant, J., Hogg, J. T., & Coltman, D. W. (2012). Genomic consequences of genetic rescue in an insular population of bighorn sheep (Ovis canadensis). *Molecular Ecology*, *21*(7), 1583–1596. https://doi.org/10.1111/j.1365-294X.2011.05427.x

Moulton, D. W., & Marshall, A. P. (2020). Laysan Duck (Anas laysanensis), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.layduc.01species_shared.bow.project_name

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology: CB*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Nolte, A. W., & Tautz, D. (2010a). Understanding the onset of hybrid speciation. *Trends in Genetics*, *26*(2), 54–58. https://doi.org/10.1016/j.tig.2009.12.001

Nolte, A. W., & Tautz, D. (2010b). Understanding the onset of hybrid speciation. *Trends in Genetics*, *26*(2), 54–58. https://doi.org/10.1016/j.tig.2009.12.001

Orozco-terwengel, P. A., & Bruford, M. W. (2014). Mixed signals from hybrid genomes. *Molecular Ecology*, *23*(16), 3941–3943. https://doi.org/10.1111/mec.12863

Ortiz, E. M. (2019). *vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis* [DOI:10.5281/zenodo.2540861]. https://zenodo.org/records/2540861/preview/edgardomortiz/vcf2phylip-v2.0.zip?include_deleted=0

Peters, J. L., McCracken, K. G., Zhuravlev, Y. N., Lu, Y., Wilson, R. E., Johnson, K. P., & Omland, K. E. (2005). Phylogenetics of wigeons and allies (Anatidae: Anas): the importance of sampling multiple loci and multiple individuals. *Molecular Phylogenetics and Evolution*, *35*(1), 209–224. https://doi.org/10.1016/j.ympev.2004.12.017

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*(3), 342–350. https://doi.org/10.1101/gr.193474.115

Reynolds, M. H., & Citta, J. J. (2007). Postfledging Survival of Laysan Ducks. *The Journal of Wildlife Management*, *71*(2), 383–388.

Reynolds, M. H., Pearce, J. M., Lavretsky, P., L, P. J., Courtot, K., & Seixas, P. P. (2015). Evidence of low genetic variation and rare alleles in a bottlenecked endangered island endemic, the Lasan Teal (Anas laysanensis). In *Technical Report* (HCSU-063; pp. 1–14). University of Hawaii at Hilo. https://pubs.usgs.gov/publication/70144555

Reynolds, M., & Klavitter, J. (2006). Translocation of wild Laysan duck Anas laysanenis to establish a population at Midway Atoll National Wildlife Refuge, United States and US Pacific Possession. *Conservation Evidence*, *3*, 6–8.

Rheindt, F. E., & Edwards, S. V. (2011). Genetic Introgression: An Integral but Neglected Component of Speciation in Birds. *The Auk*, *128*(4), 620–632. https://doi.org/10.1525/auk.2011.128.4.620

Rhymer, J. (2006). *S33-4 Extinction by hybridization and introgression in anatine ducks*. *52*(Supplement), 583–585.

Rhymer, J. M. (2006). S33-4 Extinction by hybridization and introgression in anatine ducks. *Acta Zoologica Sinica*, *52*((Supplement)), 583–585.

Rhymer, J. M., & Simberloff, D. (1996). Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), Article 4. https://doi.org/10.1038/sj.hdy.6886170

Rieseberg, L. H., & Gerber, D. (1995). Hybridization in the Catalina Island Mountain Mahogany (Cercocarpus traskiae): RAPD Evidence. *Conservation Biology*, *9*(1), 199–203.

Rieseberg, L. H., & Willis, J. H. (2007). Plant speciation. *Science (New York, N.Y.)*, *317*(5840), 910–914. https://doi.org/10.1126/science.1137729

Rigby, E. A. (2008). *Recruitment of mottled ducks (Anas fulvigula) on the upper Texas gulf coast* [Thesis]. https://ttu-ir.tdl.org/handle/2346/9189

Robinson, O. J., McGowan, C. P., & Devers, P. K. (2017). Disentangling density-dependent dynamics using full annual cycle models and Bayesian model weight updating. *Journal of Applied Ecology*, *54*(2), 670–678. https://doi.org/10.1111/1365-2664.12761

Runemark, A., Trier, C. N., Eroukhmanoff, F., Hermansen, J. S., Matschiner, M., Ravinet, M., Elgvin, T. O., & Sætre, G.-P. (2018). Variation and constraints in hybrid genome

formation. *Nature Ecology & Evolution*, *2*(3), Article 3. https://doi.org/10.1038/s41559-017-0437-7

Sæther, B.-E., Engen, S., Møller, A. P., Visser, M. E., Matthysen, E., Fiedler, W., Lambrechts, M. M., Becker, P. H., Brommer, J. E., Dickinson, J., du Feu, C., Gehlbach, F. R., Merilä, J., Rendell, W., Robertson, R. J., Thomson, D., & Török, J. (2005). Time to Extinction of Bird Populations. *Ecology*, *86*(3), 693–700.

Sætre, G.-P., & Sæther, S. A. (2010). Ecology and genetics of speciation in Ficedula flycatchers. *Molecular Ecology*, *19*(6), 1091–1106. https://doi.org/10.1111/j.1365-294X.2010.04568.x

Salazar, C. A., Jiggins, C. D., Arias, C. F., Tobler, A., Bermingham, E., & Linares, M. (2005). Hybrid incompatibility is consistent with a hybrid origin of Heliconius heurippa Hewitson from its close relatives, Heliconius cydno Doubleday and Heliconius melpomene Linnaeus. *Journal of Evolutionary Biology*, *18*(2), 247–256. https://doi.org/10.1111/j.1420-9101.2004.00839.x

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, *5*(1), 5495. https://doi.org/10.1038/ncomms6495

Schumer, M., Rosenthal, G. G., & Andolfatto, P. (2014). How Common Is Homoploid Hybrid Speciation? *Evolution*, *68*(6), 1553–1560. https://doi.org/10.1111/evo.12399

Schwarz, D., Matta, B. M., Shakir-Botteri, N. L., & McPheron, B. A. (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature*, *436*(7050), Article 7050. https://doi.org/10.1038/nature03800

Scott, A. M., Yan, J. L., Baxter, C. M., Dworkin, I., & Dukas, R. (2022). The genetic basis of variation in sexual aggression: Evolution versus social plasticity. *Molecular Ecology*, *31*(10), 2865–2881. https://doi.org/10.1111/mec.16437

Searcy, C. A., Rollins, H. B., & Shaffer, H. B. (2016). Ecological equivalency as a tool for endangered species management. *Ecological Applications*, *26*(1), 94–103. https://doi.org/10.1890/14-1674

Seehausen, O., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, *17*(1), 30–44. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Smith, G. W., & Reynolds, R. E. (1992). Hunting and Mallard Survival, 1979-88. *The Journal of Wildlife Management*, *56*(2), 306–316. https://doi.org/10.2307/3808827

Söderquist, P., Elmberg, J., Gunnarsson, G., Thulin, C.-G., Champagnon, J., Guillemain, M., Kreisinger, J., Prins, H. H. T., Crooijmans, R. P. M. A., & Kraus, R. H. S. (2017). Admixture between released and wild game birds: A changing genetic landscape in European mallards (Anas platyrhynchos). *European Journal of Wildlife Research*, *63*(6), 98. https://doi.org/10.1007/s10344-017-1156-8

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stronen, A. V., & Paquet, P. C. (2013). Perspectives on the conservation of wild hybrids. *Biological Conservation*, *167*, 390–395. https://doi.org/10.1016/j.biocon.2013.09.004

Tallmon, D. A., Luikart, G., & Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, *19*(9), 489–496. https://doi.org/10.1016/j.tree.2004.07.003

Templeton, A. R. (1986). *Coadaptation and outbreeding depression*. Sinauer Associates. https://scholar.google.com/scholar_lookup?title=Coadaptation+and+outbreeding+depression&author=Templeton%2C+A.R.&publication_year=1986

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

Trivers, R. L., & Willard, D. E. (1973). Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring. *Science*, *179*(4068), 90–92. https://doi.org/10.1126/science.179.4068.90

USFWS. (2012). *Recovery Plan for Hawaiian Waterbirds, Second Revision | U.S. Fish & Wildlife Service*. FWS.Gov. https://www.fws.gov/species-publication-action/recovery-plan-hawaiian-waterbirds-second-revision-2

Uyehara, K. J., Engilis Jr., A., & Reynolds, M. (2007). *Hawaiian Duck's Future Threatened by Feral Mallards* (Fact Sheet 3047; Fact Sheet, pp. 1–4). U.S. Fish and Wildlife Service.

van Wyk, A. M., Dalton, D. L., Hoban, S., Bruford, M. W., Russo, I.-R. M., Birss, C., Grobler, P., van Vuuren, B. J., & Kotzé, A. (2017). Quantitative evaluation of hybridization and

the impact on biodiversity conservation. *Ecology and Evolution*, *7*(1), 320–330. https://doi.org/10.1002/ece3.2595

Walters, A. D., & Schwartz, M. K. (2020). Population genomics for the management of wild vertebrate populations. *In: Hohenlohe P. A.; Rajora O. P., Eds. Population Genomics: Wildlife. Springer, Cham. p. 419-436.*, 419–436.

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution*, *35*(3), 543–548. https://doi.org/10.1093/molbev/msx319

Wells, C. P., Lavretsky, P., Sorenson, M. D., Peters, J. L., DaCosta, J. M., Turnbull, S., Uyehara, K. J., Malachowski, C. P., Dugger, B. D., Eadie, J. M., & Engilis Jr, A. (2019). Persistence of an endangered native duck, feral mallards, and multiple hybrid swarms across the main Hawaiian Islands. *Molecular Ecology*, *28*(24), 5203–5216. https://doi.org/10.1111/mec.15286

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution*, *30*(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the Risk of Extinction through Hybridization. *Conservation Biology*, *15*(4), 1039–1053.

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(W1), W52–W58. https://doi.org/10.1093/nar/gkz333

Zhang, X., Wu, R., Wang, Y., Yu, J., & Tang, H. (2020). Unzipping haplotypes in diploid and polyploid genomes. *Computational and Structural Biotechnology Journal*, *18*, 66–72. https://doi.org/10.1016/j.csbj.2019.11.011

Chapter 3: Uncovering signals of genomic islands and valleys across the speciation continuum in the Mallard complex

ABSTRACT

The continuous process of speciation as a central component of biological diversification is increasingly being appreciated, as are the resulting heterogenous signals across respective species' genomes. Consequently, resulting genomes are the product of equally dynamic evolutionary forces of genetic drift, selection, and gene flow. Although unraveling the adaptive histories of species often includes determining how directional selection uniquely acts across respective genomes, it is equally important to establish what, if any, parts remain under purifying selection for ancestral state(s). However, it is the latter that is only visible in sufficiently divergent genomes that permit the identification of both genomic islands and evolutionary valleys. Towards this, I compare the full genomes of 11 (of 14) species of the Mallard Complex that represent a successful adaptive radiation and the speciation continuum to determine how directional and purifying selection have acted to either establish derived or maintain ancestral states, respectively. First, phylogenetic and demographic (PSMC) analyses corroborated the "out-of-Africa" hypothesis for the Mallard Complex dated to 1-2 million years ago, with the most recent divergences dating to the last 500,000 years among North American species. Next, whereas genomic islands of differentiation were recovered across 44 pair-wise species comparisons representing early- and moderate-stages of divergence, genomic valleys of ancestral retention were identified across all species comparisons representing all stages of divergence. Finally, surveying gene ontology (GO) terms revealed a striking pattern in which valleys of similarity are linked to fundamental organismal functions and survival, and which outnumber islands of differentiation associated with adaptation to environmental challenges. I demonstrate

the utility in the analysis of species representing the speciation continuum to not only identify regions likely under divergent and purifying selection, but also those potential "outliers" that are in fact stochastic mirages because of genetic drift. Importantly, in the end, I find that evolutionary valleys are more predictable than genomic islands, demonstrating the strength of purifying selection in the face of speciation.

INTRODUCTION

Speciation continues to be a subject of extensive study in evolutionary biology, being recognized as a dynamic and continuous process rather than isolated event(s) (Hohenlohe et al., 2010; Nosil, Funk, et al., 2009; Nosil, Harmon, et al., 2009; Queiroz, 1998). Establishing genetic differentiation and structure between the genomes of diverging species has become standard when attempting to understand the speciation process (Andrew & Rieseberg, 2013; Ellegren et al., 2012; Irwin et al., 2018). Genomes of species are often heterogeneous because of the many evolutionary forces acting differentially across species (Andrew & Rieseberg, 2013; Feder et al., 2012; Ottenburghs et al., 2017; Turner et al., 2005; C.-I. Wu, 2001). Many recently diverged taxa often harbor much of the same genetic variability across their genomes, except for regions under strong and variable selective pressures. These selective forces contribute to the emergence of highly divergent genetic regions in a sea of similarity, a phenomenon recognized as "genomic islands of divergence" (Ellegren et al., 2012; Irwin et al., 2018; Ottenburghs et al., 2020). Theoretical and empirical studies suggest that the genomic regions that represent the magnitude of divergence between species are influenced by the extent of isolation (or gene flow) (Feder et al., 2012; Seehausen et al., 2014; C.-I. Wu, 2001). For example, parapatric and sympatric speciation, characterized by gene flow, may exhibit genetic differentiation under strong selection

and reduced recombination, while allopatric populations display heterogeneous divergence patterns across the genome, likely influenced by genetic drift and/or selection (Andrew & Rieseberg, 2013; Feder et al., 2012; Lavretsky, Dacosta, et al., 2015; Martin et al., 2013; Nosil & Feder, 2013; Yeaman & Whitlock, 2011). Consequently, genomic comparisons can unravel unique evolutionary trajectories, including differential adaptive histories.

In evolutionary biology, the concepts of purifying and directional selection stand as fundamental pillars in elucidating the dynamics of genetic variation within populations. Purifying selection, characterized by the removal of deleterious mutations from a population, serves as a crucial mechanism in maintaining the integrity of functional genes over successive generations (Cvijović et al., 2018; Stewart et al., 2008). Conversely, directional selection propels the gradual shift of allele frequencies in response to environmental pressures, favoring particular traits that enhance an organism's fitness (Brunet et al., 2021; Morales et al., 2015). Both forms of selection play distinctive roles in the intricate process of speciation. Purifying selection acts as a stabilizing force, preserving the genetic cohesion within a population, while directional selection facilitates the divergence of populations and the emergence of new species by driving adaptive evolution (Rogozin et al., 2002). These processes leave distinct genomic signatures, with directional selection resulting in regions with high genomic differentiation and purifying selection in regions of low genetic differentiation or highly conserved regions (Casillas et al., 2007; Jackson et al., 2015; Silva et al., 2014; D.-D. Wu & Zhang, 2011). A comprehensive understanding of speciation necessitates the exploration of both purifying and directional selection, as these forces collectively sculpt the genetic landscapes that define evolutionary trajectories (Brunet et al., 2021; Massingham & Goldman, 2005; Otto, 2000). Recognizing the interplay between these selective mechanisms is crucial for gaining insights into the complex

dynamics of genetic diversity and adaptation that shapes genomic diversity (Chen et al., 2017; Yang & Bielawski, 2000).

The Mallard Complex (*Anas platyrhynchos* and allies) comprises 13-14 closely related species distributed across various continents and Islands (Johnsgard, 1978; Lavretsky, McCracken, et al., 2014) that had diverged over the last million years (Johnson & Sorenson, 1999; Palmer, 1976). This radiation resulted in the Holarctic distributed mallard, while other mallard-like ducks confined to Islands or small continental ranges in the tropics or subtropics. The species complex displays diverse speciation mechanisms, including allopatry (all mallard-like ducks diverged in isolation; Lavretsky, McCracken, et al., 2014), parapatry (Mexican duck with mallard; Brown, Harrigan, et al., 2022), secondary contact (mallard with all mallard-like ducks; Lavretsky et al. 2014b), and hybrid speciation (Hawaiian duck; Lavretsky, Engilis, et al., 2015)). Being phenotypically distinct (Livezey, 1991; Palmer, 1976) but genetically similar (Lavretsky, Hernández-Baños, et al., 2014), the species within the mallard complex offer an excellent study system to study the dynamic genetic basis of speciation.

Previous studies exploring the intricacies of the speciation process have primarily focused on establishing evolutionary relationships between species, quantifying genetic differentiation, and revealing the genome-wide structure of this differentiation (Cruickshank & Hahn, 2014; Duranton et al., 2018; Irwin et al., 2018; Turner & Hahn, 2010; Wolf & Ellegren, 2017). Despite widespread acknowledgment that the majority of new mutations affecting fitness are likely deleterious (Eyre-Walker & Keightley, 2007; Jackson et al., 2015; Pál et al., 2006), the role of purifying selection in shaping large-scale differentiation patterns remains inadequately characterized. Towards this, I delve into the subtle dynamics of speciation along its continuum by assessing whole-genome sequences of 11 (of 14) species of the Mallard Complex. I provide a

comprehensive exploration of phylogenetic relationships, demographic histories, and evolutionary forces that drove diversification of this avian lineage. Importantly, my study aims to predict regions that maintain ancestral states versus those that result in divergent lineages. I hypothesize to find the same genomic regions to be under purifying selection for ancestral states despite where species are in the divergence process, whereas regions under directional selection to be less consistent.

METHODS

Sampling, DNA extraction, and whole-genome resequencing

A total of 11 muscle tissues from individuals representing 10 species and one subspecies of the Mallard Complex were collected or provided by museums (Supplementary Table S3.1), with only the extant eastern spot-billed duck (*A. zonorhyncha*), Indian spot-billed duck (*A. poecilorhyncha*), and Meller's duck (*A. melleri*), as well as the extinct Mariana mallard (*Anas oustaleti*) missing. Note that preliminary population genetics analyses based on partial-genome data and against sets of respective reference samples of each taxa ensured that all samples had ≥ 98% assignment probability to their respective taxonomic lineages; ensuring that none of my specimens were contemporary hybrids or harbored genetic diversity estimates that were not representative of their taxa. For eight samples, DNA was extracted using QIAGEN Blood & Cell Culture DNA kit following the manufacturer's protocols (Qiagen). Extracted DNA was cleaned using the Clean and Concentrated^{TM} -25 kit following the manufacturer's protocols (Zymo Research). Following, DNA concentration was determined using Qubit fluorometer v 2.0 (Life Technologies), and DNA quality was determined using NanoDrop 2000 spectrophotometer (Thermo Scientific). Clean DNA was sent to Novogenetics LTD. (Sacramento, California) for

sequencing on an Illumina HiSeq X with 150PE chemistry at a minimum of 30Gb per sample. Finally, tissue for the remaining three species (i.e. Laysan duck, Hawaiian duck, and mallard) were sent to Cantata Bio, LLC (Scotts Valley, CA) where DNA was extracted, followed by proprietary Chicago library preparation, and sequencing on an Illumina HiSeq X with 150PE chemistry at a minimum of 30 Gb per sample (also see (Lavretsky, Hernández, et al., 2023).

Bioinformatics

All raw Illumina reads were filtered and aligned using a custom in-house Python script (Python scripts available at https://github.com/jonmohl/PopGen; Lavretsky et al.2020). Note that raw published sequences for the Tufted duck (*Aythya fuligula;* retrieved from NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_009819795.1/; Kraus et al. Unpublished) the Greater Scaup (*Aythya marila*; retrieved from NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_029042245.1/; Zhou, Unpublished), and the Baer's pochard (*Aythya baeri*; retrieved from NCBI, https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_026413565.1/; Zhang, Unpublished) were included in alignments, but only further processed in downstream phylogenetic analyses where they served as outgroups. I first merged the Illumina paired-end sequences with the program PEAR v.0.9.11 (Zhang et al., 2014), then I trimmed and discarded poor quality reads using Trimmomatic (Bolger et al., 2014). All quality sequences were then aligned to the wild North American mallard reference genome (Lavretsky et al., 2023; Available at: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030704485.1/) using the Burrows Wheeler Aligner v07.15 (bwa; Li and Durbin, 2009). Samples were then sorted and indexed in Samtools v1.6 (Li et al., 2009) and combined and genotyped using the bcftools v1.6 "mpileup" and "call"

functions with the following parameters "-c –A -Q 30 -q 30", which set a base pair and overall sequence PHRED score of ≥30 to ensure that high-quality sequences only are retained. Finally, alleles were further filtered for a minimum allele depth < 5 (--remove-filtered 'AD<5'), and genotypes were required to have a minimum depth of 10 (--minDP 10) from the resulting VCF file using VCFtools v0.1.17 (Danecek et al., 2011).

Summary statistics

Estimates of pair-wise species relative differentiation (F_{ST}), absolute divergence (D_{XY}), and per species nucleotide diversity (π) were calculated based on non-overlapping 100-Kb windows size using the custom script popgenWindows.py script from Martin et al. (2015) (Available in: https://github.com/simonhmartin/genomics_general.). Additionally, runs of homozygosity (ROH) were also calculated for each genome using all possible base-pairs as implemented in VCFtools v0.1.17 (Danecek et al., 2011).

Phylogenetic analysis

I first filtered the full VCF file that included outgroups to include single nucleotide polymorphisms (SNPs) that were 100 bp apart using VCFtools v0.1.17 (i.e., --thin 100; Danecek et al., 2011). The SNP dataset was further filtered with bcftools v1.6 (i.e., bcftools view-i 'COUNT(GT= "RR") > 0 & COUNT(GT= "AA") > 0'; Li et al., 2009) to retain only those SNPs present at least once in my dataset as homozygous for the reference allele and homozygous for the alternative allele. I then partitioned datasets that included autosomal or the Z-sex chromosome only before generating alignments in FASTA file format using vcf2phylip.py (Ortiz, 2019). For each of these alignments I estimated the best nucleotide substitution model

using IQ-Tree (i.e., --s; Nguyen et al., 2015). I then reconstructed Maximum Likelihood species trees with 100 bootstrap replicates for each autosomal or Z-sex chromosome linked loci in RAxML-NG v.1.1.0 (Kozlov et al., 2019). Resulting trees were plotted and visualized in Figtree v.1.4.4 (Rambaut, 2012).

Divergence times were estimated for the autosomal based tree only using the penalized likelihood approach as implemented in treePL v 1.0 (Sanderson, 2002; Smith & O'Meara, 2012). The adjusted calibration points used to estimate divergence time were retrieved from timetree (Kumar et al. 2022), and included *Anas platyrhynchos – Aythya marila* (14.4 Mya; Mitchel et al 2014, Jiang et al. 2010, Jiang et al. 2014, Roquet et al 2014), and *Aythya fuligula – Aythya marila* (2.3 Mya; Fulton et al. 2012, Roquet et al. 2014). Note that I conducted a priming analysis to determine the best optimization parameters before running 18 cross-validation analyses separated by 10-fold increments starting from $1x10^{-12}$ to find the best smoothing value. The resulting tree was visualized in FigTree v1.4.4 (Rambaut, 2012).

Reconstructing demographic histories

I used the Pairwise sequential Markovian Coalescent (PSMC) model (Li & Durbin, 2011) to independently reconstruct the demographic histories from whole-genome sequences for each of the 11 taxa. PSMC estimates rates of coalescent events across a single genome and uses these to infer effective population size (N_E) in the past. PSMC analysis requires a consensus genome sequence (FASTQ) generated using SAMtools v1.3.1 (Li et al., 2009), bcftools v 1.1 (Narasimhan et al., 2016), and the vcftutils.pl scrip from bcftools to call variants from the alignment with the following command "samtools mpileup -C50 -uf ref.fas aln.bam| bcftools call -c | vcfutils.pl vcf2fq -d 10 -D 100 > diploid.fq". The minimum depth (d) and maximum depth

(D) were set up to approximately a third and twice the average depth read of each genome, respectively, and as recommended the PSMC documentation (https://github.com/lh3/psmc). PSMC parameters were all set as N = 30, t = 5, r = 5 and p ="4+30*2+4+6+10", and run with 100 bootstrap replicates (Nadachowska-Brzyska et al., 2015). Time and effective population sizes were transformed into biologically relevant numbers using generation time (G) and mutation rate (i.e., $1.0x10^{-9}$ per site per generation; Lavretsky et al. 2020). Note that the generation time was based on each species' average maximum age at sexual maturity (α) and adult survival (s) (i.e., $G = \alpha + (s / (1 - s))$; see Supplementary Table S3.2).

Identifying genes within high and low divergence areas

Relative differentiation (F_{ST}) was estimated across 55 pair-wise species comparisons and in 100-kb windows. Note that my study was interested in determining the repeatability in the number of locations of genomic regions that are under selection, with a more specific goal, of understanding how these differ between regions under directional versus purifying selection. As a result, I did not consider any regions specific to a single taxon, but those that were found across multiple. To do so, genomic regions putatively under directional or purifying selection were based on windows representing the highest or lowest 1%, respectively, in each of the 55 comparisons. I then counted the times a selected window appeared across all 55 comparisons. Windows with at least half of the highest count per chromosome were considered (Table 3.1). For example, if an autosomal window in the highest 1% F_{ST} appeared in a maximum of 10 pair-wise comparisons, then I also included any region appearing in a minimum of five comparisons. However, I also considered windows that appeared in fewer comparisons if they were located between selected windows found in the minimum number of comparisons as these regions could

indicate areas under genetic hitchhiking (a.k.a. genetic draft; (Neher, 2013), and may indicate genes of evolutionary importance. Finally, I used in-house Python scripts to extract genes within windows meeting my threshold criteria to investigate for enrichment of Gene Ontology(GO) terms, including calculating significance with a Bonferroni correction at an adjusted *p*-value cutoff of 0.05.

Table 3.1. Counting method of windows per chromosome.

Chromosome	Number of times a window appeared across all 55 comparisons for the highest 1% FST	# Windows meeting the threshold (half of the highest count)	Number of times a window appeared across all 55 comparisons for the lowest 1% FST	# Windows meeting the threshold (half of the highest count)
1	21	10	55	27
2	19	9	48	24
3	9	4	49	24
4	12	6	38	19
5	12	6	47	23
6	16	8	29	14
7	16	8	48	24
8	12	6	36	18
9	9	4	43	21
10	12	6	54	27
11	16	8	28	14
12	4	2	36	18
13	13	6	46	23
14	8	4	31	15
15	7	3	39	19
16	5	2	39	19
17	1	0	54	27
18	10	5	55	27
19	10	5	53	26
20	9	4	30	15
21	7	3	55	27

22	6	3	52	26
23	2	1	18	9
24	10	5	46	23
25	3	1	25	12
26	4	2	35	17
27	7	3	32	16
28	7	3	55	27
29	8	4	49	24
Z	9	4	46	23

RESULTS

Sequencing results

Total Illumina sequencing data ranged from 17 – 90x mean depth coverage genomes (Supplementary Table S3.3). Alignment of all genomes resulted in a total of 884,965,881bps (of ~1.04 Gbp) of overlapping sequence, and which contained 64,588,080 single nucleotide polymorphism (SNPs) across the 11 duck taxa and the outgroup (Supplementary Table S3.3).

Summary statistics

Estimating relative differentiation (F_{ST}) across pair-wise species comparisons yielded varying levels of genomic divergence that represent the speciation continuum (Figure 3.1). In general, values were remarkably high between the Laysan duck and the remaining species (Avg. F_{ST} range = 0.31 – 0.71), followed by the African Black duck (*Anas sparsa*) with a similar pattern (Figure 3.2A). Genomes with high levels of relative differentiation tended to carry lower levels of nucleotide diversity (Avg. π = 0.00085) and longer runs of homozygosity (Figure 3.2 B,C). Interestingly, the African Black duck had one of the lowest levels of nucleotide diversity (Avg. π = 0.0036), but also had short runs of homozygosity (Avg. ROH = 6,388bp). Conversely, species representing the North American clade – i.e., mallard, the American black duck, the Mexican

duck and both Mottled ducks – had the lowest levels of relative genomic differentiation (Avg. F_{ST} range = 0.05 – 0.42) (Figure 3.2A), but the highest levels of nucleotide diversity (Avg. π range = 0.007 – 0.008), and variable runs of homozygosity (Avg. ROH range = 6,996 bp – 20,552 bp) (Figure 3.2C).

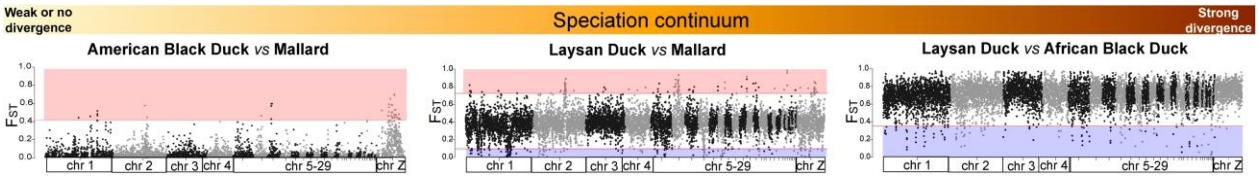

Figure 3.1. Speciation continuum: The continuous nature of divergence during speciation reflected in the relative differentiation (F_{ST}) in pairwise comparison from low genetic differentiation (left) to high genetic differentiation (right) along the whole-genome sequences.

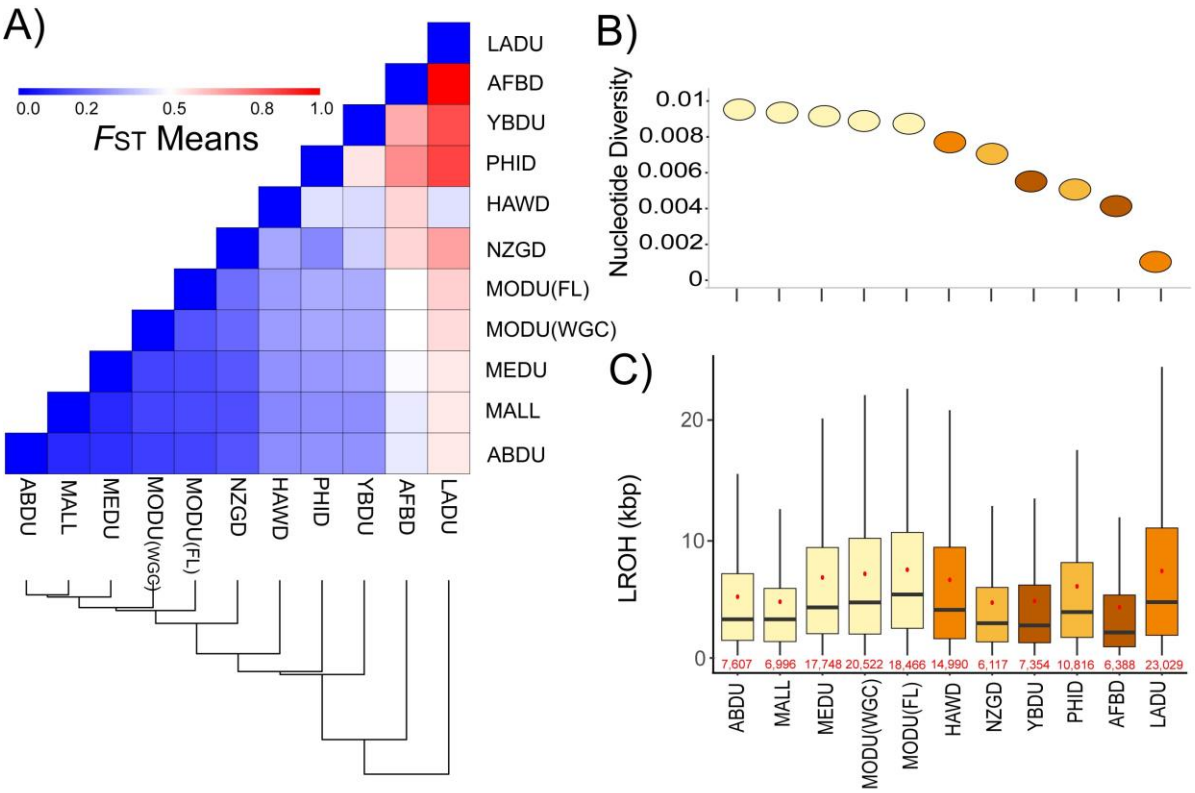

Figure 3.2. Summary statistics: A) Heatmap and relatedness based on means F_{ST}, B) Means of nucleotide diversity, and C) Boxplots of run of homozygosity (ROH) per species into the Mallard complex. Note that red dots along with their values denote average ROH in each plot. Color scale is determined by their phylogenetic relationship.

Phylogenetic analysis

A total of 345,299 quality base pairs met my filtering criteria when including outgroups, with 320,023 bps (109,404 SNPs) and 24,826 bps (8,0878 SNPs) for the autosomal and Z-sex chromosome datasets, respectively. Bootstrap support values across phylogenetic nodes were high, with slightly higher values for the Z-sex chromosome compared to the autosomal based phylogeny. Regardless, four main taxonomic groups were recovered: (1) the North American clade, (2) Hawaiian Island clade, (3) Australasian Pacific clade, and (4) the African clade (Figure 3.3). First, the most basal group was the African clade in both trees. Conversely, I found that Hawaiian Island and Australasian Pacific clades swapping positions, being sister to the North American clade in Autosomal and the Z-sex chromosome-based trees, respectively (Figure 3.3). Finally, whereas the mallard was found as sister to the Mexican duck in the autosomal based tree, the mallard placed as the outgroup to the remaining North American ducks in the Z-sex chromosome-based trees.

Estimating nodal divergences in the autosomal tree suggest that the deepest split in the Mallard Complex occurred in the late Pliocene around 2.78 million years ago (Mya), while most species-level divergences occurred during the early Pleistocene between 0.72 to 1.73 Mya (Figure 3.2A). More specifically, the African clade showed the oldest split at 2.27 Mya, while the North American clade had the most recent divergence times, placing species evolution to the early Pleistocene between 0.72 to 1.30 Mya.

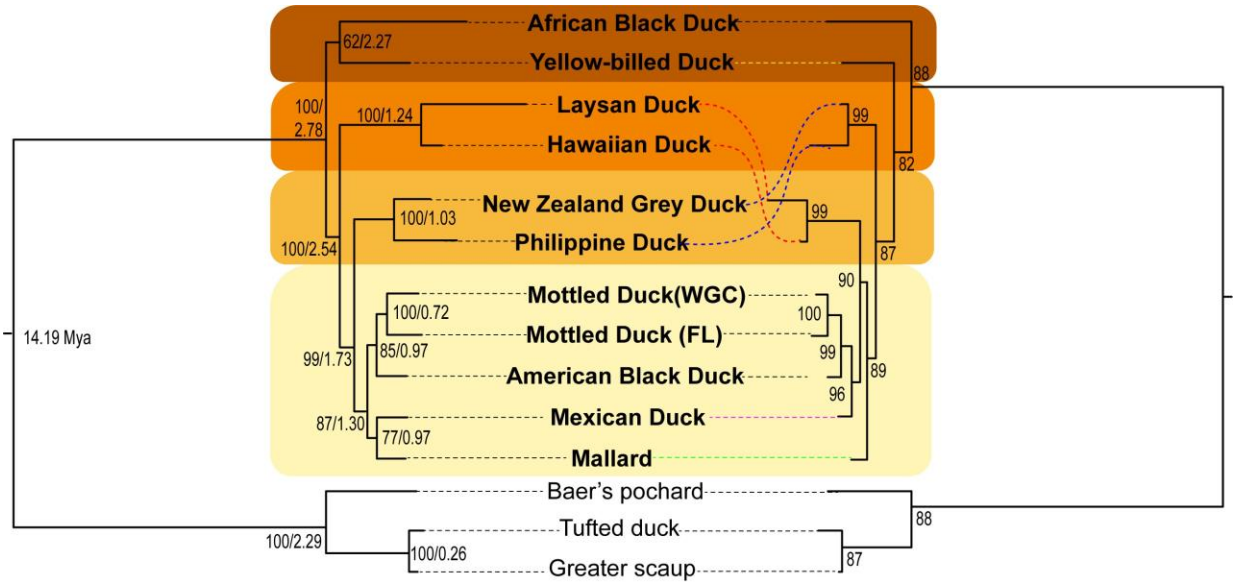

Figure 3.3: Phylogenetic tree: (Left) Based on autosomes and (Right) Based on the Z-sex chromosome.

Demographic history

Using the PSMC method, I was able to trace the fluctuations in effective population size across the 11 taxa, covering most of the Pleistocene period, from as early as 10,000 years ago to as far back as nine million years (Figure 3.4). First, notable shifts in respective N_E were observed in the Mallard, Hawaiian duck, Philippine duck, Laysan duck, and Florida (FL) Mottled duck dating to the early Pleistocene cooling period (3-0.9 Mya) (Figure 3.4 B, C,I,J,K), with each of these experiencing increasing N_E but displaying distinct trajectories by the end of this period. The Philippine duck and the FL Mottled duck experienced rising N_E during the mid-Brunhes event (MBE, 430-110 kya), followed by a steady decline until 10,000 years ago. Similarly, the mallard mirrored this pattern, but its N_E stabilized at the beginning of the Last Ice age (LIA, 110 kya). Conversely, both the Laysan duck and the Hawaiian duck exhibited a sharp decrease in effective population size during the LIA, despite a slight increase at the end of the MBE. The remining duck species, including the Yellow-billed duck, (WGC) Mottled duck, Mexican duck,

American Black duck, African Black duck, and New Zealand Grey Duck, showed a steady increase in their N_E at the beginning of the MBE, continuing into the beginning of the LIA, but with different trajectories by the Last Glacial Maximum (LGM, 26,500-19,000) (Figure 3.4 A,D, E,F,G,H). During this last glacial period, the Yellow-billed duck, Mexican duck, and New Zealand Grey duck exhibited an increase in their N_E, followed by a decline by the period's end, except for the New Zealand Grey duck that continued to increase. Conversely, the African Black duck, WGC Mottled duck, and American Black duck experienced a decrease in effective population size during the LGM, with the WGC Mottled duck and African black duck continuing a constant decline afterward, while the American Black duck showed a trend toward increasing its population size.

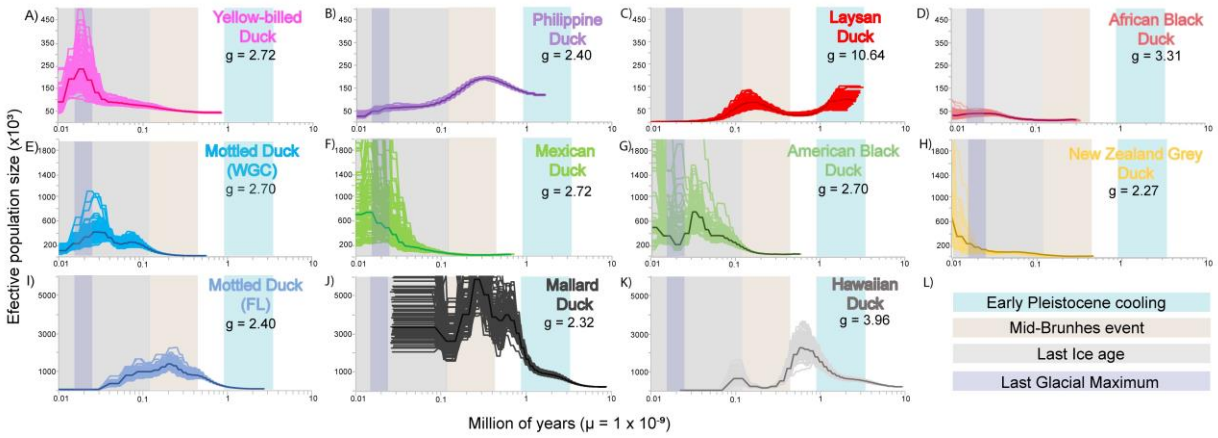

Figure 3.4: Demographic analyses: PSMC results of each mallard-like duck.

Identifying genes within high and low divergence areas

Outlier windows were based on relative differentiation (F_{ST}) estimated across all 55 pair-wise species comparisons and based on non-overlapping 100 kbp windows. For autosomes, the maximum number of pair-wise comparisons for any of the top and bottom 1% of F_{ST} windows was 21 and 55, respectively, resulting in the selection of those windows found in a minimum of

10 and 25 comparisons for respective top and bottom windows. For the Z-sex chromosome, the maximum count for the top and bottom 1% of F_{ST} windows was 9 and 46, respectively, resulting in the selection of those windows found in a minimum of 4 and 23 comparisons for respective top and bottom windows (Table 3.1). Windows within the highest 1% of F_{ST} estimates across autosomes harbored a total of 194 gene ontology terms (GO terms), 47 of which were found to be significantly enriched (Figure 3.5, Table S3.4, S3.5). Similarly, a total of 360 GO terms were identified in the windows within the lowest 1% F_{ST} estimates of autosomes, but only 17 were found to be significantly enriched (Figure 3.5, Table S3.4, S3.6). Conversely, windows within the highest 1% of F_{ST} estimates across the Z-sex chromosome harbored a total of 130 GO terms, 17 of which were significantly enriched, whereas a total of 107 GO terms were found in the windows within the lowest 1% F_{ST}, with 57 being significantly enriched (Figure 3.5, Table S3.4, S3.5, S3.6). Interestingly, prior to enrichment analyses, I recovered most GO terms among the lowest and highest F_{ST} windows for Autosomes and the Z-sex chromosome, respectively (Figure 3.5). However, this relationship flips when considering statistical significance in enrichment, with most enriched GO terms being in the highest versus the lowest F_{ST} windows for Autosomes and the Z-sex chromosome, respectively (Figure 3.5). Additionally, I recovered ~10% of GO terms shared among the highest and lowest F_{ST} windows, with only ~3% of these being significantly enriched (Figure 3.5).

GO terms that were significantly enriched among autosomes were associated with positive regulation in several processes related to immune systems, reproduction, and growth (Table S3.4, S3.5). Among GO terms that were significantly enriched within the Z-sex chromosome, I identified processes related to respiratory and circulatory system function, as well as inflammatory response and reproduction (Table S3.4, S3.6). Finally, a total of 51 and 27 GO

terms were found to be in both highest and lowest F_{ST} windows in autosomes and the Z-sex chromosome, respectively, but with only two of these being significantly enriched in both instances (Figure 3.5), and associated with reverse transcription (i.e., GO:0006278-"RNA-directed DNA polymerase activity and GO:0003964-"RNA-dependent DNA biosynthetic process").

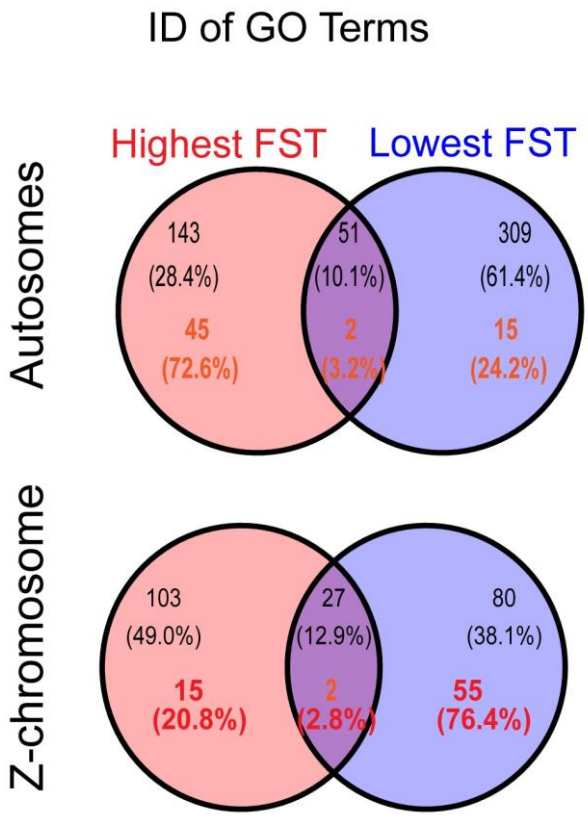

Figure 3.5: Ven Diagram: GO Term enrichment analyses for Autosome genes and Z-sex genes for the 1% highest and lowest F_{ST}. The number in red represents significant number of Go Terms.

DISCUSSION

Here, I provide the first genomic analysis of taxa representing a recently radiated Mallard Complex. In general, whole-genome data provided higher nodal support than before, but with taxonomic relationships following those derived from either multiple loci (Lavretsky,

McCracken, et al., 2014) or partial-genome data (Lavretsky, DaCosta, et al., 2019). Once again, I provide support for the "out-of-Africa" hypothesis for the Mallard Complex (Lavretsky, McCracken, et al., 2014; Palmer, 1976), with the African clade being basal to the others in both autosomal and Z-sex chromosomes-based trees (Figure 3.3). However, whereas the Hawaiian clade is sister to the North American clade in the Autosomal based tree (bootstrap support = 100), it is the Oceanic clade that is sister to the North American clade in the Z-sex chromosomes-based tree (bootstrap support = 87). Previously, Lavretsky et al (2014) hypothesized that the dispersal into the Holarctic and Oceania likely proceeded simultaneously, and thus, clade switching may be because of varying evolutionary histories of these marker-types during the divergence process (Degnan & Rosenberg, 2009; Ellegren et al., 2012; Peters et al., 2014). Moreover, this clade switching may simply be because of the unique hybrid evolutionary history of Hawaiian ducks, causing signal biases in bifurcating tree analyses that do not capture more complex evolutionary histories (i.e., hybrid speciation; Lavretsky et al., 2015b). The only other discordance between the autosomal or Z-sex chromosome-based trees is the placement of the mallard that is either sister to the Mexican duck or outgroup to the remaining North American species, respectively. The same discordance in marker-type based tree reconstruction was recovered when analyzing multiple individuals and partial-genome data; however, mallards were sister to Mexican ducks when analyzing non-outlier autosomal regions only (Lavretsky, DaCosta, et al., 2019). In fact, the topology of the North American clade derived in my Z-sex chromosome-based tree was identical to trees derived from double digest restriction-site associated DNA sequencing (ddRADseq) SNPs of Autosomal outlier regions or the Z-sex chromosome regardless if they were recovered in putatively outlier or non-outlier regions. Together, I conclude that recency of the North American clades divergence resulting in high

rates of incomplete-lineage sorting (ILS) likely biases autosomal based trees when including all possible sites, moving the mallard to the species that most recently diverged from it (i.e., Mexican duck; Lavretsky et al., 2019).

The topology of my Z-sex chromosome-based tree (Figure 3.3) corresponds with previous studies more specifically examining evolutionary histories of Mexican ducks (Brown et al., 2022), American black ducks (Lavretsky, Janzen, et al., 2019), and mottled ducks (Peters et al., 2016) that all support each of the four North American monochromatic species diverging nearly simultaneously from the dichromatic mallard in the last 500,000 years. Indeed, species-specific demographic analyses generally recover non-zero effective population sizes for American black ducks, Mexican ducks, and WGC mottled ducks starting only 100,000 years ago, which are all distinct from when the non-zero effective population size of wild mallards dating to 1.5 million years ago (Figure 3.4). I posit that genetic drift may be biasing demographic inferences for FL mottled ducks that are the only ones with near-zero demographics today, suggesting these likely have gone through a severe bottleneck more recently; which is known to distort demographic analyses (Brown et al., 2022; Lavretsky et al., 2023). Unlike the FL mottled duck, the American black duck (Figure 3.4G) and Mexican duck (Figure 3.4F) have experienced general population growths with today's effective population size being ~600,000. Whereas the WGC mottled duck experienced population declines since ~ 30,000 years ago, their effective population remains at 100,000 today (Figure 3.4E). Together, evolutionary models support a scenario in which the mallard that evolved in Eurasia millions of years ago (Lavretsky, Mohl, et al., 2023) invaded and expanded in North America in the last million years, before pockets of them were isolated during glaciation events that resulted in the loss of dichromatism and evolution of these ecological

replacements post-glaciation and over the last 500,000 years (Lavretsky, DaCosta, et al., 2019; Omland, 1997).

Next, my phylogenetic and demographic analyses place the evolution of this species complex to the last 2.5 to 1 million years before present, respectively (Figures 3.3 & 3.4). Nevertheless, the disparity in time estimates between methods was primarily driven by those derived from my autosomal based phylogenetic tree that were nearly twice that of demographic analyses and previous studies (Brown et al., 2022; Lavretsky et al., 2019a, 2019b, 2015a, 2015b, 2014b). This elevated time estimate is likely a result of the dating method in treePL, which incorporates age constraints based on both fossil data and ages reported in source papers (Smith & O'Meara, 2012; Tims et al., 2021). In my study, the dating method relied solely on previous reported age estimates in papers (see methods) because of the lack of fossil or geological evidence within the Anatidae group. Assessing divergence time estimation using molecular data is crucial in biology, however several challenges remain including nucleotide substitution rates and fossil calibration (Pennington et al., 2004; Püschel et al., 2020). I stress the importance of considering fossil to date the Anatidae group whenever possible to mitigate the disagreements in age estimates.

Evolutionary histories and standing genetic diversity within a species radiation

Although I found a general negatively correlated trend between nucleotide diversity and runs of homozygosity (ROH) as would be expected (Bosse et al., 2019; Curik et al., 2014; Willoughby et al., 2015), my result was not significant (*p-value* = 0.718; Supplementary Figure S3.1). Increases in ROH proceeded with declining nucleotide diversity except within the two subspecies of mottled ducks and the Mexican duck (Figure 3.2B, 3.2C). Specifically, despite

having similar mean nucleotide diversity as the other North American clade taxa (Figure 3.2B), and which were also concordant with previous estimates (Lavretsky, DaCosta, et al., 2019; Peters et al., 2016), they had the highest average ROH of any taxa (Figure 3.2C). In fact, the only other species with similar ROH estimates is the Laysan duck, which is known to have gone through a severe bottleneck to 11 individuals in 1911 (Reynolds & Klavitter, 2006). This event explains its having the lowest levels of nucleotide diversity (Figure 3.2B) and corresponding to previous population genetics studies uncovering limited genetic diversity at several nuclear introns and major histocompatibility genes (Lavretsky, Engilis Jr, et al., 2014; Lavretsky, McCracken, et al., 2014). This discordance in mottled ducks and the Mexican duck suggest specific demographic or evolutionary scenarios, such as recent population bottleneck, founder effect or inbreeding within small populations. During such events, genetic diversity is reduced because of loss of rare alleles, leading to increase in homozygosity (Allendorf et al., 2001; Bosse et al., 2019; Hedrick, 2015). However, during the subsequent expansion phases or ongoing recombination events, new mutations may arise, leading to increases in nucleotide diversity that is unequally distributed across the genome. Indeed, previous studies have suggested that mottled duck populations reflect historical isolation and divergence by neutral genetic drift (Peters et al. 2016), and Mexican ducks evolving through sequential founder events (Brown et al 2022) both potentially contributing to their long runs of homozygosity. Similarly, the African Black duck shows disparities between nucleotide diversity and ROH, harboring low levels of nucleotide diversity but short runs of homozygosity (Figure 3.2B, 3.2C). Although the cause for these disparities will require additional genomes and analyses, it is evident that nucleotide diversity is not always equally distributed across the genome because of specific historical, demographic, or evolutionary factors influencing the genetic landscape of populations.

The genetic diversity of each taxon expectedly translated to unique demographic histories (Figure 3.4; Kim et al., 2016; Nadachowska-Brzyska et al., 2015). First, mallards and American black ducks that harbored the highest mean calculated diversity (Figure 3.2B) and shortest ROH (Figure 3.2C) also had some of the most exponential growths, along with cycles in effective population size in their respective time-sets (Figure 3.4G & 3.4J). Next, yellow-billed ducks and New Zealand grey ducks represent species with moderate levels of nucleotide diversity and ROH generally (Fogure 3.2B&3.2C), and each of these also have exponentially increasing effective population sizes (Figure 3.4A&3.4H) but more gradually than mallards or American black ducks. Finally, Philippine ducks and Laysan ducks are species that represent the tail-end of both nucleotide diversity and ROH estimations that suggests strong effects of genetic drift in their evolutionary histories, and coinciding with recovered demographic histories (i.e., Philippine duck & Laysan duck; Figures 3.4B & 3.4C). However, severe cases of genetic drift are known to strongly bias demographic analyses (Walsh et al., 2019), and thus, I caution interpretating demographic analyses for these species. Although, the yellow-billed duck's demographic history extends one million years before present and may represent the extant ancestor of this species complex, the lack of diversity in the African black duck's genome limited resolution (Figure 3.4D); and thus, determining which of these may have been the most basal lineage remains unresolved. Similarly, the demographic history reconstructed for the Hawaiian duck requires careful consideration as they represent a putative hybrid species (Lavretsky, Engilis, et al., 2015). In fact, the recovered pattern of effective population size for Hawaiian ducks (Figure 3.4K) is indeed intermediate in pattern between those recovered in their putative parental taxa, mallards (Figure 3.4J) and Laysan ducks (Figure 3.4C). Together, I conclude that researchers attempting

to understand evolutionary histories through similar demographic analyses need to carefully consider whether their genomes may have been affected by severe bottlenecking or represent ancient or contemporary hybrids, as these factors can introduce significant biases (Brown, Harrigan, et al., 2022; Mattingsdal et al., 2020).

Genomic comparisons across the speciation continuum reveals selective heterogeneity

Estimating pair-wise species relative differentiation (F_{ST}) revealed genomes representing the speciation continuum (Figure 3.1). Whereas species at the earliest stages of divergence are important to identify genomic regions putatively under directional selection (i.e., American black duck vs Mallard), I found that those at the intermediate and later stages of divergence provided a view of genomic valleys where purifying selection maintains ancestral polymorphisms (i.e., Laysan duck vs African black duck; Figure 3.1). Traditionally, evolutionary studies focused on identifying genomics islands responsible for species divergence (Duranton et al., 2018; Ottenburghs et al., 2020; Renaut et al., 2013; Sendell-Price et al., 2020; Turner et al., 2005; Wolf & Ellegren, 2017). However, I argue that understanding the strength and location of purifying selection across species complexes is equally important (Cvijović et al., 2018; Hofer et al., 2012; Roesti et al., 2012; Sendell-Price et al., 2020; Van Doren et al., 2017; Wang et al., 2016). This perspective provides clues on how ancestral states are maintained in the face of genetic drift and divergent selection. To do so, I apply a novel method to demarcate the 1% F_{ST} outlier regions putatively under directional (i.e. islands) versus purifying (i.e., valleys) selection, which resulted in 98 and 196 total regions, respectively (Supplementary Table S3.4). Specifically, a total of 63 and 182 windows were recovered as top and bottom 1% in autosomal chromosomes, while 35 and 14 windows represented the top and bottom 1% in the Z-sex chromosome. Notably, previous

studies have used extreme percentiles of the distribution (with an arbitrary threshold) for a given summary statistics, such as F_{ST}, to identify outliers from divergent peaks (Campagna et al., 2017; Walsh et al., 2019; Wolf & Ellegren, 2017), or weighted statistical techniques, such as kernel smoother to identify both high or low diverged regions (Sendell-Price et al., 2020; Van Doren et al., 2017). My results align with the expected divergence process, wherein differentiation accumulates between populations, initially forming distinct regions or islands of differentiation at the beginning of the species continuum. As populations progress along this continuum, genomic islands expand because of linkage disequilibrium, facilitating divergence of neutral and weakly selected loci via divergence hitchhiking (Nosil et al. 2009, Feder and Nosil 2010). Towards the middle of the species continuum, highly conserved regions (i.e. genomic valleys) emerge, characterized by significantly lower differentiation compared to the background levels, potentially influencing genomic heterogeneity and slowing genome-wide divergence. My findings underscore the pivotal roles of genomic islands and valleys in shaping the genomic landscape across the speciation continuum, evident across all mallard-like duck population comparisons.

In general, the functional patterns of Gene Ontology (GO) Terms were related with both adaptation and reproductive isolation as evidenced by their distribution across autosomes and the Z-sex chromosome. Those identified in the top 1% in autosomes and Z-sex chromosome were linked to potential directional selection, likely playing crucial roles in evolutionary adaptation or diversification (Supplementary Table S3.3, S3.5). Many of these GO terms were closely associated in reproductive processes, such as the developmental process involved reproduction (GO:0003006), oocyte maturation (GO:0001556), as well as positive regulation of oocyte

development (GO:0060282), all of which are crucial for sex determination and development (Machado et al., 2022; Tang et al., 2021). Such mechanisms play a fundamental role in increasing reproductive isolation (Stöck et al., 2021). Moreover, the enrichment analysis suggests association with pathways involved in immune response (GO:0002684, GO:0070233), growth regulation (GO:0003420, GO:1902733, GO:0008083), development processes (GO:0007431, GO:0008340), or hormone activity (GO:0005179), allowing organisms to adapt to different environmental challenges or pathogen pressures. Conversely, significant GO terms in the bottom 1% that are associated with potential purifying selection, enrichment analysis suggests tight connections to essential metabolic pathways or highly conserved protein-coding regions (Supplementary Table S3.3, S3.6). For example, glutamate-5-semialdehyde dehydrogenase activity (GO:0004350) is related to molecular functions such as biosynthesis of amino acids from the glutamine family, such as arginine, glutamate, gutamine and proline (QuickGo,2024). Genes related to amino acid biosynthesis are often conserved (Heizer et al., 2011), highlighting their crucial role in cellular function and survival. Similarly, enriched GO terms found in the Z-sex chromosome are also involved in essential cellular processes crucial for organismal function and survival. These processes include DNA replication (GO:0006166), RNA transcription (GO:0042789), protein synthesis (GO:0019509), cellular signaling (GO:0017061, GO:0018344, GO:1902165), and intracellular transport (GO:0016021, GO:0031422). Finally, I argue that any GO terms in the windows comprising both islands and valleys are more likely a result of random drift, as these clearly can be fixed or lost in these species (Figure 3.5). This likely explains why only ~2% of these GO terms were recovered as statistically enriched. Together, these comparisons provide a means to better understand the roles

of directional and purifying selection, as well as random chance (i.e., genetic drift) play in shaping genomic diversity.

REFERENCES

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Andrew, R. L., & Rieseberg, L. H. (2013). Divergence is focused on few genomic regions early in speciation: Incipient speciation of sunflower ecotypes. *Evolution; International Journal of Organic Evolution*, *67*(9), 2468–2482. https://doi.org/10.1111/evo.12106

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bosse, M., Megens, H.-J., Derks, M. F. L., de Cara, Á. M. R., & Groenen, M. A. M. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary Applications*, *12*(1), 6–17. https://doi.org/10.1111/eva.12691

Brown, J. I., Harrigan, R. J., & Lavretsky, P. (2022). Evolutionary and ecological drivers of local adaptation and speciation in a North American avian species complex. *Molecular Ecology*, *31*(9), 2578–2593. https://doi.org/10.1111/mec.16423

Brown, J. I., Hernández, F., Engilis, A., Hernández-Baños, B. E., Collins, D., & Lavretsky, P. (2022). *Genomic and morphological data shed light on the complexities of shared ancestry between closely related duck species*. *12*(1), 10212. https://doi.org/10.1038/s41598-022-14270-2

Brunet, T. D. P., Doolittle, W. F., & Bielawski, J. P. (2021). The role of purifying selection in the origin and maintenance of complex function. *Studies in History and Philosophy of Science Part A*, *87*, 125–135. https://doi.org/10.1016/j.shpsa.2021.03.005

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3*(5), e1602404. https://doi.org/10.1126/sciadv.1602404

Casillas, S., Barbadilla, A., & Bergman, C. M. (2007). Purifying Selection Maintains Highly Conserved Noncoding Sequences in Drosophila. *Molecular Biology and Evolution*, *24*(10), 2222–2234. https://doi.org/10.1093/molbev/msm150

Chen, J., Glémin, S., & Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution*, *34*(6), 1417–1428. https://doi.org/10.1093/molbev/msx088

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics*, *209*(4), 1235–1278. https://doi.org/10.1534/genetics.118.301058

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant

call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution*, *24*(6), 332–340. https://doi.org/10.1016/j.tree.2009.01.009

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04963-6

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, *491*(7426), Article 7426. https://doi.org/10.1038/nature11584

Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews. Genetics*, *8*(8), 610–618. https://doi.org/10.1038/nrg2146

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics: TIG*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Hedrick, P. W. (2015). Heterozygote Advantage: The Effect of Artificial Selection in Livestock and Pets. *Journal of Heredity*, *106*(2), 141–154. https://doi.org/10.1093/jhered/esu070

Heizer, E. M., Raymer, M. L., & Krane, D. E. (2011). Amino Acid Biosynthetic Cost and Protein Conservation. *Journal of Molecular Evolution*, *72*(5), 466–473. https://doi.org/10.1007/s00239-011-9445-4

Hofer, T., Foll, M., & Excoffier, L. (2012). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*, *13*, 107. https://doi.org/10.1186/1471-2164-13-107

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862. https://doi.org/10.1371/journal.pgen.1000862

Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, *27*(23), 4839–4855. https://doi.org/10.1111/mec.14858

Jackson, B. C., Campos, J. L., & Zeng, K. (2015). The effects of purifying selection on patterns of genetic differentiation between Drosophila melanogaster populations. *Heredity*, *114*(2), Article 2. https://doi.org/10.1038/hdy.2014.80

Johnsgard, P. A. (1978). Ducks, geese, and swans of the world. *(No Title)*. https://cir.nii.ac.jp/crid/1130282272525895680

Johnson, K. P., & Sorenson, M. D. (1999). Phylogeny and Biogeography of Dabbling Ducks (Genus: Anas): A Comparison of Molecular and Morphological Evidence. *The Auk*, *116*(3), 792–805. https://doi.org/10.2307/4089339

Kim, S., Cho, Y. S., Kim, H.-M., Chung, O., Kim, H., Jho, S., Seomun, H., Kim, J., Bang, W. Y., Kim, C., An, J., Bae, C. H., Bhak, Y., Jeon, S., Yoon, H., Kim, Y., Jun, J., Lee, H., Cho, S., … Yeo, J.-H. (2016). Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome Biology*, *17*(1), 211. https://doi.org/10.1186/s13059-016-1071-4

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Lavretsky, P., Dacosta, J. M., Hernández-Baños, B. E., Engilis, A., Sorenson, M. D., & Peters, J. L. (2015). Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology*, *24*(21), 5364–5378. https://doi.org/10.1111/mec.13402

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*(10), 2594–2609. https://doi.org/10.1111/mec.15091

Lavretsky, P., Engilis, A., Eadie, J. M., & Peters, J. L. (2015). Genetic admixture supports an ancient hybrid origin of the endangered Hawaiian duck. *Journal of Evolutionary Biology*, *28*(5), 1005–1015. https://doi.org/10.1111/jeb.12637

Lavretsky, P., Engilis Jr, A., & Peters, J. L. (2014). Major histocompatibility I gene diversity in the critically endangered Laysan duck (Anas laysanensis). *Pacific Conservation Biology*, *20*(1), 86–93.

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. E. (2023). Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos). *G3 Genes|Genomes|Genetics*, *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., Hernández-Baños, B. E., & Peters, J. L. (2014). Rapid radiation and hybridization contribute to weak differentiation and hinder phylogenetic inferences in the New World

Mallard complex (Anas spp.). *The Auk*, *131*(4), 524–538. https://doi.org/10.1642/AUK-13-164.1

Lavretsky, P., Janzen, T., & McCracken, K. G. (2019). Identifying hybrids & the genomics of hybridization: Mallards & American black ducks of Eastern North America. *Ecology and Evolution*, *9*(6), 3470–3490. https://doi.org/10.1002/ece3.4981

Lavretsky, P., McCracken, K. G., & Peters, J. L. (2014). Phylogenetics of a recent radiation in the mallards and allies (Aves: Anas): Inferences from a genomic transect and the multispecies coalescent. *Molecular Phylogenetics and Evolution*, *70*, 402–411. https://doi.org/10.1016/j.ympev.2013.08.008

Lavretsky, P., Mohl, J. E., Söderquist, P., Kraus, R. H. S., Schummer, M. L., & Brown, J. I. (2023). The meaning of wild: Genetic and adaptive consequences from large-scale releases of domestic mallards. *Communications Biology*, *6*(1), 1–15. https://doi.org/10.1038/s42003-023-05170-w

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), Article 7357. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Livezey, B. C. (1991). A Phylogenetic Analysis and Classification of Recent Dabbling Ducks (Tribe Anatini) Based on Comparative Morphology. *The Auk*, *108*(3), 471–507. https://doi.org/10.2307/4088089

Machado, A. M., Fernández-Boo, S., Nande, M., Pinto, R., Costas, B., & Castro, L. F. C. (2022). The male and female gonad transcriptome of the edible sea urchin, *Paracentrotus lividus*: Identification of sex-related and lipid biosynthesis genes. *Aquaculture Reports*, *22*, 100936. https://doi.org/10.1016/j.aqrep.2021.100936

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Research*, *23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Massingham, T., & Goldman, N. (2005). Detecting Amino Acid Sites Under Positive Selection and Purifying Selection. *Genetics*, *169*(3), 1753–1762. https://doi.org/10.1534/genetics.104.032144

Mattingsdal, M., Jorde, P. E., Knutsen, H., Jentoft, S., Stenseth, N. C., Sodeland, M., Robalo, J. I., Hansen, M. M., André, C., & Blanco Gonzalez, E. (2020). Demographic history has shaped the strongly differentiated corkwing wrasse populations in Northern Europe. *Molecular Ecology*, *29*(1), 160–171. https://doi.org/10.1111/mec.15310

Morales, H. E., Pavlova, A., Joseph, L., & Sunnucks, P. (2015). Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Molecular Ecology, 24*(11), 2820–2837. https://doi.org/10.1111/mec.13203

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology: CB, 25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics, 32*(11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Neher, R. A. (2013). Genetic Draft, Selective Interference, and Population Genetics of Rapid Adaptation. *Annual Review of Ecology, Evolution, and Systematics, 44*(Volume 44, 2013), 195–215. https://doi.org/10.1146/annurev-ecolsys-110512-135920

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution, 32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nosil, P., & Feder, J. L. (2013). Genome evolution and speciation: Toward quantitative descriptions of pattern and process. *Evolution; International Journal of Organic Evolution, 67*(9), 2461–2467. https://doi.org/10.1111/evo.12191

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*(3), 145–156. https://doi.org/10.1016/j.tree.2008.10.011

Omland, K. E. (1997). Examining Two Standard Assumptions of Ancestral Reconstructions: Repeated Loss of Dichromatism in Dabbling Ducks (anatini). *Evolution*, *51*(5), 1636–1646. https://doi.org/10.1111/j.1558-5646.1997.tb01486.x

Ortiz, E. M. (2019). *vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis* [DOI:10.5281/zenodo.2540861]. https://zenodo.org/records/2540861/preview/edgardomortiz/vcf2phylip-v2.0.zip?include_deleted=0

Ottenburghs, J., Honka, J., Müskens, G. J. D. M., & Ellegren, H. (2020). Recent introgression between Taiga Bean Goose and Tundra Bean Goose results in a largely homogeneous landscape of genetic differentiation. *Heredity*, *125*(1), Article 1. https://doi.org/10.1038/s41437-020-0322-z

Ottenburghs, J., Megens, H.-J., Kraus, R. H. S., van Hooft, P., van Wieren, S. E., Crooijmans, R. P. M. A., Ydenberg, R. C., Groenen, M. A. M., & Prins, H. H. T. (2017). A history of hybrids? Genomic patterns of introgression in the True Geese. *BMC Evolutionary Biology*, *17*(1), 201. https://doi.org/10.1186/s12862-017-1048-2

Otto, S. P. (2000). Detecting the form of selection from DNA sequence data. *Trends in Genetics*, *16*(12), 526–529. https://doi.org/10.1016/S0168-9525(00)02141-7

Pál, C., Papp, B., & Lercher, M. J. (2006). An integrated view of protein evolution. *Nature Reviews. Genetics*, *7*(5), 337–348. https://doi.org/10.1038/nrg1838

Palmer, R. S. (Ed.). (1976). *Handbook of North American Birds Volume II: Waterfowl*.

Pennington, P. T., Cronk, Q. C. B., Richardson, J. A., Near, T. J., & Sanderson, M. J. (2004). Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil–based model selection. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1450), 1477–1483. https://doi.org/10.1098/rstb.2004.1523

Peters, J. L., Lavretsky, P., DaCosta, J. M., Bielefeld, R. R., Feddersen, J. C., & Sorenson, M. D. (2016). Population genomic data delineate conservation units in mottled ducks (Anas fulvigula). *Biological Conservation*, *203*, 272–281.

Peters, J. L., Winker, K., Millam, K. C., Lavretsky, P., Kulikova, I., Wilson, R. E., Zhuravlev, Y. N., & McCracken, K. G. (2014). Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus Anas). *Molecular Ecology*, *23*(12), 2961–2974. https://doi.org/10.1111/mec.12799

Püschel, H. P., O'Reilly, J. E., Pisani, D., & Donoghue, P. C. J. (2020). The impact of fossil stratigraphic ranges on tip-calibration, and the accuracy and precision of divergence time estimates. *Palaeontology*, *63*(1), 67–83. https://doi.org/10.1111/pala.12443

Queiroz, K. de. (1998). *The general lineage concept of species, species criteria, and the process of speciation*. https://www.academia.edu/87407560/The_general_lineage_concept_of_species_species_criteria_and_the_process_of_speciation

Rambaut, A. (2012). *FigTree*. http://tree.bio.ed.ac.uk/software/figtree/

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., & Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, *4*, 1827. https://doi.org/10.1038/ncomms2833

Reynolds, M., & Klavitter, J. (2006). Translocation of wild Laysan duck Anas laysanenis to establish a population at Midway Atoll National Wildlife Refuge, United States and US Pacific Possession. *Conservation Evidence*, *3*, 6–8.

Roesti, M., Hendry, A. P., Salzburger, W., & Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, *21*(12), 2852–2862. https://doi.org/10.1111/j.1365-294X.2012.05509.x

Rogozin, I. B., Spiridonov, A. N., Sorokin, A. V., Wolf, Y. I., Jordan, I. K., Tatusov, R. L., & Koonin, E. V. (2002). Purifying and directional selection in overlapping prokaryotic genes. *Trends in Genetics*, *18*(5), 228–232. https://doi.org/10.1016/S0168-9525(02)02649-5

Sanderson, M. J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. *Molecular Biology and Evolution*, *19*(1), 101–109. https://doi.org/10.1093/oxfordjournals.molbev.a003974

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews. Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., Coulson, T., & Clegg, S. M. (2020). The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (Zosterops lateralis). *G3 Genes|Genomes|Genetics*, *10*(9), 3147–3163. https://doi.org/10.1534/g3.120.401352

Silva, D. R. D., Nichols, R., & Elgar, G. (2014). Purifying Selection in Deeply Conserved Human Enhancers Is More Consistent than in Coding Sequences. *PLOS ONE*, *9*(7), e103357. https://doi.org/10.1371/journal.pone.0103357

Smith, S. A., & O'Meara, B. C. (2012). treePL: Divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Stewart, J. B., Freyer, C., Elson, J. L., Wredenberg, A., Cansu, Z., Trifunovic, A., & Larsson, N.-G. (2008). Strong Purifying Selection in Transmission of Mammalian Mitochondrial DNA. *PLOS Biology*, *6*(1), e10. https://doi.org/10.1371/journal.pbio.0060010

Stöck, M., Dedukh, D., Reifová, R., Lamatsch, D. K., Starostová, Z., & Janko, K. (2021). Sex chromosomes in meiotic, hemiclonal, clonal and polyploid hybrid vertebrates: Along the 'extended speciation continuum'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1833), 20200103. https://doi.org/10.1098/rstb.2020.0103

Tang, Y., Chen, J.-Y., Ding, G.-H., & Lin, Z.-H. (2021). Analyzing the gonadal transcriptome of the frog Hoplobatrachus rugulosus to identify genes involved in sex development. *BMC Genomics*, *22*(1), 552. https://doi.org/10.1186/s12864-021-07879-6

Tims, A. R., Unmack, P. J., & Ho, S. Y. W. (2021). A fossil-calibrated time-tree of all Australian freshwater fishes. *Molecular Phylogenetics and Evolution*, *161*, 107180. https://doi.org/10.1016/j.ympev.2021.107180

Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, *19*(5), 848–850. https://doi.org/10.1111/j.1365-294X.2010.04532.x

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, *26*(15), 3982–3997. https://doi.org/10.1111/mec.14083

Walsh, J., Clucas, G. V., MacManes, M. D., Thomas, W. K., & Kovach, A. I. (2019). Divergent selection and drift shape the genomes of two avian sister species spanning a saline–freshwater ecotone. *Ecology and Evolution*, *9*(23), 13477–13494. https://doi.org/10.1002/ece3.5804

Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. K. (2016). Variation in Linked Selection and Recombination Drive Genomic Divergence during Allopatric Speciation of European and American Aspens. *Molecular Biology and Evolution*, *33*(7), 1754–1767. https://doi.org/10.1093/molbev/msw051

Willoughby, J. R., Fernandez, N. B., Lamb, M. C., Ivy, J. A., Lacy, R. C., & DeWoody, J. A. (2015). The impacts of inbreeding, drift and selection on genetic diversity in captive

breeding populations. *Molecular Ecology*, *24*(1), 98–110. https://doi.org/10.1111/mec.13020

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), Article 2. https://doi.org/10.1038/nrg.2016.133

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

Wu, D.-D., & Zhang, Y.-P. (2011). Different level of population differentiation among human genes. *BMC Evolutionary Biology*, *11*(1), 16. https://doi.org/10.1186/1471-2148-11-16

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, *15*(12), 496–503. https://doi.org/10.1016/S0169-5347(00)01994-7

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution; International Journal of Organic Evolution*, *65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Chapter 4: Unknown Biases: Understanding how anthropogenic hybridization impact genomic diversity and demographic estimates

ABSTRACT

Anthropogenic hybridization can profoundly influence the genetic landscape of wild populations. In particular, whereas the crossing of two outbred lineages may result in increasing genetic diversity, though likely resulting in lost local adaptations, hybridization with inbred individuals can result in decreasing diversity. Here, I focus on the latter as I investigate consequences of hybridization between domestic and wild mallard-like ducks across diverse ecological regions and with varied stocking histories in Hawaii, New Zealand, and North America. I re-sequence whole-genomes representing domestic and wild congeners, as well as the resulting hybrids that span admixture levels across ecoregions. For each genome, I assessed genetic diversity, runs of homozygosity, and estimated demographic parameters. First, I report that wild genomes consistently harbored higher levels of genetic diversity, lower runs of homozygosity, and lower inbreeding coefficients as compared to the respective domestic lineage. Importantly, the resulting summary statistics for the respective admixed individuals followed proportional contributions of their parental groups, including hybrids tending towards domestic ancestry possessing higher inbreeding coefficients. Among ecoregions, however, I recovered the strongest statistical correlation between ancestry and summary statistics for North America where domestic game-farm mallards continue to be released and annually interact with wild populations. Conversely, both Islands in which stocking efforts largely ceased decades ago and/or populations are maintained at low numbers I find evidence of adaptive selection and genetic drift as the overwhelming forces on the genomes in New Zealand and Hawaii, respectively. Finally, I

demonstrate how inferences of demographic histories using genomes born of the domestication process or hybrid origins can be extremely biased, resulting in distorted estimates of effective population size (Ne) and divergence times. I discuss the intricate interaction among domestication, anthropogenic hybridization, and wildlife conservation in the Anthropocene. Together, I conclude that the use of highly inbred individuals when attempting to supplement populations of wild congeners not only results in high levels of hybridization, but that the resulting offspring are increasingly inbred themselves, which has real conservation and adaptive consequences.

INTRODUCTION

Anthropogenic hybridization refers to the process whereby human activities, directly or indirectly, lead to interbreeding between previously isolated taxa (McFarlane & Pemberton, 2019; Ottenburghs, 2021). The frequency of such events are increases because of human-mediated habitat modifications, fragmentation, and the introduction of exotic species that facilitate the breakdown of reproductive isolation between sympatric species (Grabenstein & Taylor, 2018) or result in secondary contact between previously allopatric taxa (McFarlane & Pemberton, 2019). Anthropogenic hybridization can have significant constructive or destructive outcomes for the interbreeding populations. It can be beneficial in terms of adaptive introgression or an increase in genetic diversity (Crispo et al., 2011; Ottenburghs, 2021). However, it can also result in genetic swamping or extinction by hybridization (Rhymer & Simberloff, 1996; Todesco et al., 2016). In recent decades, studies have highlighted numerous examples of anthropogenic hybridization occurring across a range of taxa, from plants (Guo, 2014; Lamont et al., 2003) to mammals (Adavoudi & Pilot, 2022; Salvatori et al., 2019) and

birds (Casas et al., 2012; Grabenstein et al., 2023; Lavretsky, Mohl, et al., 2023), recognizing the evolutionary significance of hybridization. In conservation biology, anthropogenic hybridization poses challenges for species management and conservation efforts, as it can threaten their genetic native status or loss of local adaptations (Allendorf et al., 2001; Rhymer & Simberloff, 1996). Understanding the mechanisms and consequences of anthropogenic hybridization is therefore important for developing effective strategies to mitigate its impacts on Biodiversity.

An increasing concern within the realm of anthropogenic hybridization involves the interbreeding between domesticated animals and their wild relatives. The uncontrolled diffusion of domestic organisms, including establishment of feral populations, might lead to loss of genetic diversity and eventual local extinctions via introgressive hybridization (Randi, 2008; Rhymer & Simberloff, 1996). In particular, such interbreeding may not only introduce genes favored under artificial selection that are maladaptive in natural environments, but can also result in the disruption of locally adaptive gene complexes; and both leading to a reduction in the fitness of wild populations (Kidd et al., 2009). Examples of interbreeding between domesticated and wild counterparts includes terrestrial carnivores (dogs, cats), ungulates (pigs, goats), game stocks (Galliformes, waterfowl), and many fish species (Ottenburghs, 2021; Rhymer & Simberloff, 1996).

Finally, introgressed domestic genes can result in complex genetic patterns that may obscure demographic signals and bias population genetic analyses (Brown et al., 2022; Lavretsky, Hernández, et al., 2023; Lu et al., 2022). Despite the advances of genome-sequencing and demographic models such as Pairwise Sequentially Markovian Coalescent (PSMC, Li and Durbin, 2011), few studies have explored how introgressive hybridization may affect PSMC inferences. In particular, analyzing genomes of hybrids that vary in admixed ancestry is required

to fully realize the power of PSMC when attempting to reconstruct population histories from confounded genomes.

Study System

The world-wide popularity of the mallard (*Anas platyrhynchos*) among hunters and consumers has led to the generation of various domestic breeds, including captive breeding of strains specifically aimed for release into the wild to increase recreational opportunities (Lavretsky, Mohl, et al., 2023; Schummer et al., 2023; Söderquist et al., 2017). However, the consequences of these human activities on the resilience of wild populations, particularly in relation to hybridization with these domestic mallards in their natural habitats, remain largely unknown. In particular, the release of the domestic game-farm mallard strain has resulted in wide-spread hybridization in all instances, including the formation of hybrid swarm of the Hawaiian duck (*Anas wyviliana*) and mallard outside the island of Kaua'i (Wells et al. 2019) and the New Zealand grey duck (*Anas supercilious*; Brown, 2021). Briefly, game-farm mallards were first imported to the Hawaiian Islands in early 1800s for food and hunting purposes (Engilis, Pyle, Davis 2004, Pyle and Pyle 2017). The Hawaiian duck was extirpated from the main Hawaiian Islands except Kaua'i by the 1960s, and while reintroductions were conducted in O'ahu, Maui, and Hawai'i from the late 1950s through the early 1990s (USFWS, 2012), the presence of feral mallards on those islands resulted in feral mallard x Hawaiian duck hybrid swarms by the 2000s (Wells et al. 2019). Similarly, the release of over 30,000 game-farm mallards into New Zealand between the 1860s to the 1950s resulted in a current census population of ~3 million (Williams, 2017) and resulting in wide-spread hybridization with and near genetic extinction of the endemic New Zealand grey duck (Brown, 2021). Whereas these endemic Island species were always found in small numbers (Engilis Jr. et al., 2020; Gillespie,

1985; USGS, 2007), which may explain their vulnerability to genetic extinction, a similar scenario is occurring at a more continental scales in Eurasia (Champagnon et al., 2013; Söderquist et al., 2017) and North America (Lavretsky, Mohl, et al., 2023), where game-farm mallards continue to be released in mass today. In North America, the release of game-farm mallards across the last 100 years has recently been confirmed to have resulted in widespread and continuous hybridization with wild mallards. The eastern part of North American is now considered a game-farm x wild mallards hybrid swarm (Lavretsky et al., 2020; Lavretsky, Mohl, et al., 2023). Whereas these examples highlight the profound impact of human activities can have on wild populations, they also present an opportunity to test whether hybridization results in similar consequences as each instance is between the same domestic lineage (Lavretsky et al. 2023).

Here, I aim to explore the consequences of secondary contact between domestic and wild congeners on genomic diversity and demographic estimates using regions that exhibit diverse ecologies and differing degrees of domesticated mallard release, including Hawaii, New Zealand, and North America. I analyzed whole-genome sequences of genetically vetted domestic and wild parental species, as well as various backcrosses that comprise identified domestic x wild swarms in the three regions. The domestication process often profoundly alters the genomic landscape towards increasingly lowered levels of genetic diversity as these breeds are exposed to serial generations of artificial selection; whether intended or not (Kidd et al., 2009; Ottenburghs, 2021). As a result, I predict that genomes representing domestic lineages will have lower levels of genetic diversity, increased runs of homozygosity, and generally higher levels of inbreeding as compared to their local wild relative (Crispo et al., 2011; Hedrick, 2015). Next, I hypothesize that the same measures of diversity/inbreeding will scale with the domestic level of admixture

among hybrid offspring. Finally, studies have demonstrated that genomes that have undergone severe bottlenecking or are highly admixed can result in biased demographic inferences (C. A. Martin et al., 2023). Thus, I also take this opportunity to determine levels of bias when estimating effective population size (N_E), often a critical value used to assess conservation status (Lu et al., 2022), using genomes of domestic or hybrid origin. Together, my study provides insight into the genetic consequences of anthropogenic hybridization and its impact on overall genetic diversity of wild populations; something that is increasingly needed in the Anthropocene (Buckley & Catford, 2016; Ottenburghs, 2021).

METHODS

Sampling, DNA extraction, whole-genome resequencing, and variant calling

A total of 18 muscle tissue samples representing wild and domestic duck populations found in the Hawaiian Islands, New Zealand, and North America were collected or provided by museums (Supplementary Table S4.1). Note that all samples were picked by previous ancestry assignments obtained across population genetics studies using partial-genome data conducted in Hawaii (Wells et al. 2019), New Zealand (Brown 2021), and North America (Lavretsky 2023). Doing so, increased the likelihood each sample represented each genetic group recovered in those studies. Next, two (of 18) samples (i.e. Hawaiian duck, and North American wild mallard) were first sent to Cantata Bio, LLC (Scotts Valley, CA) where DNA was extracted, followed by proprietary Chicago library preparation, and sequencing on an Illumina HiSeq X to ~330M read pairs using 150bp paired-end (PE) chemistry per sample. For the remaining 16 tissues, DNA was extracted using QIAGEN Blood & Cell Culture DNA kit according to the manufacturer's protocols (Qiagen), followed by additional cleaning using the Clean and Concentrated^{TM} -25 kit

as per manufacturer's instructions (Zymo Research). Subsequently, DNA concentration and quality were determined using Qubit fluorometer v 2.0 (Life Technologies),and a NanoDrop spectrophotometer (Nanodrop 2000, Thermo Scientific), respectively. Extracted DNA was then sent to Novogenetics LTD (Sacramento, California) for library preparation and sequencing using 150 bp PE chemistry on a NovaSeq 6000 until ≥30Gb was achieved per sample.

Raw Illumina paired-end reads were first merged with the program PEAR v.0.9.11 (Zhang et al., 2014), then I trimmed or discarded poor quality sequences using Trimmomatic (Bolger et al., 2014), followed by sequence alignment to the recently published reference-scale wild North American mallard genome (Lavretsky et al., 2023; Available at: https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_030704485.1/) using the Burrows Wheeler Aligner v. 07.15 (bwa; Li and Durbin, 2009). Samples were then sorted and indexed in Samtools v1.6 (H. Li et al., 2009) and combined and genotyped using the bcftools "mpileup" and "call" functions with the following parameters "-c –A -Q 30 -q 30", which set a base pair and overall sequence PHRED score of ≥30 to ensure that high-quality sequences only are retained. The resulting VCF file was then filtered using VCFTOOLS v.0.1.17 (Danecek et al., 2011) with a minimum quality of 30 (-minQ 30), a minimum depth of 10 (-minDP 10) and removing all sites with a minimum allele depth of 5 (-remove-filtered 'AD<5'). Note that bioinformatic steps were automated using the in-house Python script (Python scripts available at https://github.com/jonmohl/PopGen; Lavretsky et al.2020).

Measures of genetic diversity

For each genome, I assessed genetic diversity by calculating nucleotide diversity (π) across non-overlapping 100-Kb windows using the custom script popgenWindows.py script from Martin et

al. (2015) (Available in: https://github.com/simonhmartin/genomics_general.). For runs of homozygosity (ROH), I used VCFtools v0.1.17 (Danecek et al., 2011) and calculated using all possible base-pairs. Finally, SNP-based inbreeding coefficients (F_{IS}) were calculated using VCFtools v0.1.17 (Danecek et al., 2011).

Inference of demographic history

I inferred effective population size (N_E) across time for each genome using the Pairwise sequential Markovian Coalescent (PSMC) model (H. Li & Durbin, 2011). PSMC analyses require a consensus genome sequence (FASTQ) generated using SAMtools v1.3.1 (H. Li et al., 2009), bcftools v1.1 (Narasimhan et al., 2016), and the vcftutils.pl scrip from bcftools to call variants from the alignment with the following command "samtools mpileup -C50 -uf ref.fas aln.bam| bcftools call -c | vcfutils.pl vcf2fq -d 10 -D 100 > diploid.fq". The minimum depth (d) and maximum depth (D) were set up to approximately a third and twice the average depth read of each genome, respectively, and as recommended (see https://github.com/lh3/psmc). Parameters were set to include N= 30, t =5, r= 5 and p ="4+30*2+4+6+10", and analyses run with 100 bootstrap replicates (Nadachowska-Brzyska et al., 2015). The generation time (G) was calculated for each duck based on their average maximum age at sexual maturity (α) and adult survival (s) following the following formula G = = $\alpha + (s / (1 - s))$, and using a mutation rate of to 1.0 $x10^{-9}$ per site per generation (Lavretsky et al. 2020). For mallard-like ducks, the age of maturity is typically 1 year (i.e. $\alpha = 1$; Alerstam & Högstedt, 1982). However, I estimated the average survival rate individually. For wild mallards, this survival rate is ~0.57 (i.e. range: 0.46–0.68; (Arnold & Clark, 1996; Drilling et al., 2020; Smith & Reynolds, 1992); resulting in an estimated generation time of 2.32 years. For New Zealand grey duck, the survival rate is 0.56

(Caithness et al., 1991), leading to an estimated generation time of 2.27 years. For Hawaiian ducks, the survival rate is 0.74 (Malakowski, in review), resulting in an estimated generation time of 3.97 years. For all the domestic mallards and the domestic x wild hybrids, I used the same estimated generation time as for wild mallards.

RESULTS

All genomes were sequenced to mean coverage depths ranging from 17x – 84x (Table S4.2), with high alignment similarity (~99%) across the 879,691,675 quality base pairs, including 63,805,135 single nucleotide polymorphism (SNPs) that passed filtering. Note that I limited downstream analyses to autosomes only.

Measures of genetic diversity

Among genomes representing wild species, mallards had the highest nucleotide diversity (Avg. π = 0.0083), smallest ROH (Avg. Autosomal ROH = 5,306 bp), and lowest inbreeding coefficients (Avg. F_{IS} = 0.23) as compared to Hawaiian ducks (Avg. π = 0.0068, Avg. Autosomal ROH = 24,642 bp, Avg. F_{IS} = 0.35) and New Zealand grey ducks (Avg. π = 0.0063, Avg. Autosomal ROH = 5,582 bp, Avg. F_{IS} = 0.39) (Table 4.1). In each case, however, summary statistics for the representative domestic genomes were not necessarily worse as compared to their respective endemic wild congener, with those from Hawaii (Avg. π = 0.0067, Avg. Autosomal ROH = 37,389 bp, Avg. F_{IS} = 0.37) generally having less favorable statistics compared to New Zealand (Avg. π = 0.0074, Avg. Autosomal ROH = 15,255 bp, Avg. F_{IS} = 0.28) and North America (Avg. π = 0.0068, Avg. Autosomal ROH = 26,991 bp, Avg. F_{IS} = 0.34). In fact, genomes representing game-farm x wild mallard hybrids in North America were the only set that showed expected and

statistically significant correlations between levels of domestic ancestry and summary statistics, with ROH and inbreeding coefficients being positively and nucleotide diversity negatively correlated (Figure 4.1A). Although correlation between levels of domestic ancestry and summary statistics were also statistically significant for genomes from New Zealand, I found increasing nucleotide diversity and decreasing inbreeding coefficients with increasing domestic ancestry; although ROH remained positively correlated with domestic ancestry, the values were generally very short (Figure 4.1B; Table 4.1). Finally, none of the summary statistics were significantly correlated with domestic ancestry among genomes from Hawaii, with generally similar summary statistics regardless of domestic ancestry levels (Figure 4.1C; Table 4.1).

Table 4.1. Per sample average nucleotide diversity (π), runs of homozygosity (ROH), and inbreeding coefficients (F_{IS}), as well as the effective population size (N_E) obtained from the most recent time estimate from demographic analyses (Figs. 3-4). Note that samples are broken up by ecoregion and includes representative category based on the proportion of domestic ancestry assigned from partial genome data from previous population genetics studies (Wells et al. 2019, Brown 2021, Lavretsky et al. 2023).

Ecoregion	Category	Domestic Parent Ancestry	Wild Parent Ancestry	F	π	ROH	Ne
North America	Wild mallard	0.02	0.98	0.2268	0.0083	5306	2,900,000
	Hybrid 1	0.20	0.80	0.2033	0.0083	5939	4,780,000
	Hybrid 2	0.21	0.79	0.2094	0.0083	7378	320,000
	Hybrid 3	0.61	0.39	0.2647	0.0076	17444	35,000
	Game-farm mallard	1.00	0.00	0.3435	0.0068	26991	15,000
New Zealand	New Zealand Grey Duck	0.00	1.00	0.3927	0.0063	5582	700,000
	Hybrid 1	0.09	0.91	0.4061	0.0061	5982	280,000
	Hybrid 2	0.21	0.79	0.3314	0.0069	5359	1,070,000
	Hybrid 3	0.50	0.50	0.2688	0.0075	6271	1,890,000
	Hybrid 4	0.76	0.24	0.2769	0.0075	9396	880,000
	New Zealand mallard	1.00	0.00	0.2804	0.0074	15255	230,000
Hawaii	Hawaiian duck	0.00	1.00	0.3636	0.0068	14734	50,000
	Hybrid 1	0.19	0.81	0.4835	0.0053	39071	147,000
	Hybrid 2	0.53	0.47	0.3514	0.0067	30548	23,000

Hybrid 3	0.81	0.19	0.3743	0.0064	37389	200,000
Park mallard	1.00	0.00	0.3499	0.0067	24642	16,000

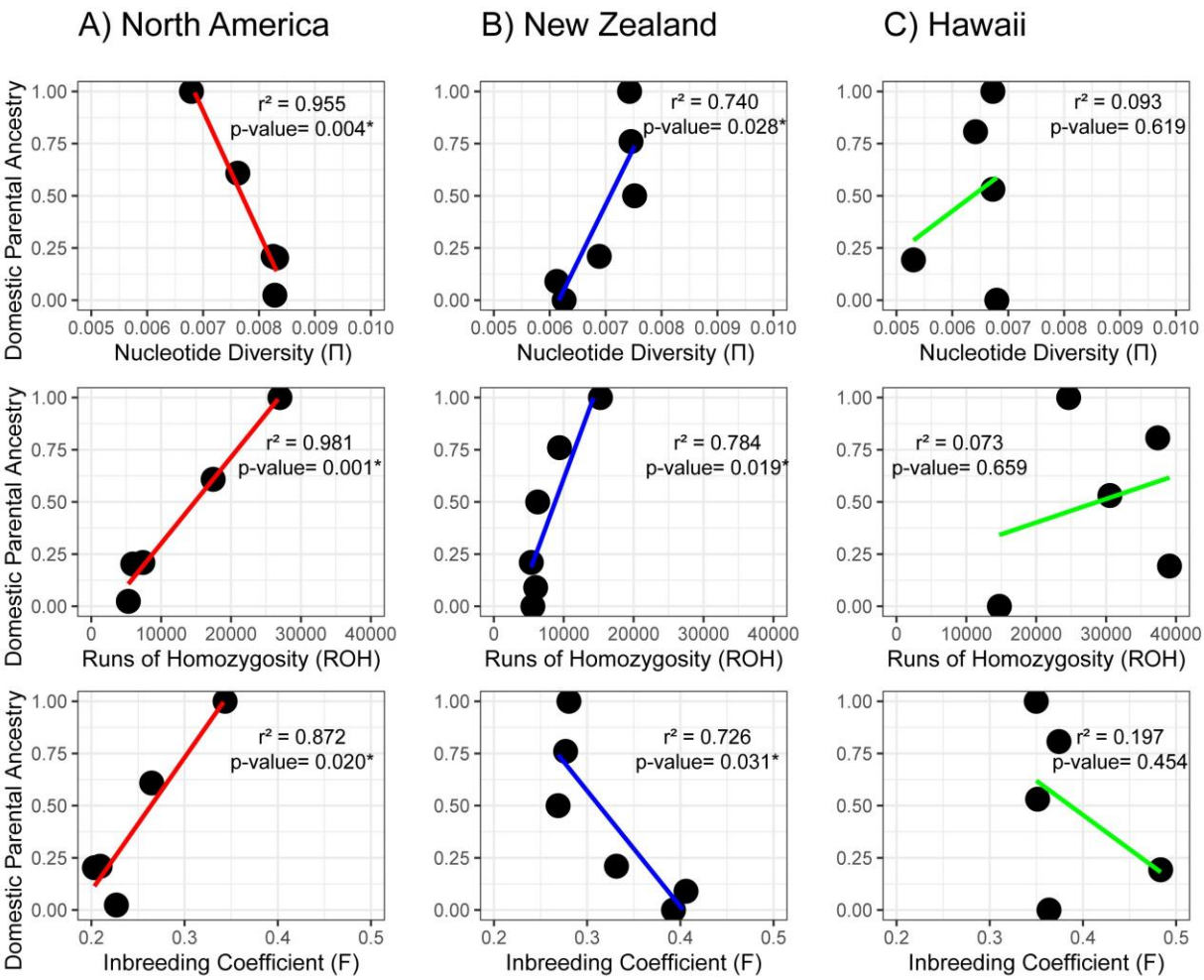

Figure 4.1: Correlation between domestic parental ancestry and measures of genetic diversity per sample.

Demographic history

PSMC analyses reliably estimated effective population size (N_E) up to at least 1.5 million years before present (Figures 4.2 – 4.4). Among genomes representing wild species, both the wild North American mallard and the Hawaiian duck exhibit similar demographic patterns, with an increasing effective population size starting around 1 million years before present (YBP). The

wild North American mallard reached its peak effective population size around 175,000 YBP before declining, stabilizing at an effective population size of ~3 million since 100,000 YBP. In contrast, the Hawaiian duck reached its peak effective population size around 570,000 YBP before experiencing declines, followed by a second increase around 100,000 YBP, and stabilizing at a contemporary effective population size of around 50,000 individuals. Finally, I recovered a continuous increase in effective population size since around 150,000 YBP for the New Zealand grey duck, reaching a contemporary effective population size of 100,000.

Next, the general trend in the increase of N_E in deep time followed by a continuous reduction to near-zero today that was recovered across genomes representing domestic lineages is concordant with the fact that both Hawaii's and New Zealand's feral populations and those that are being released in North America are genetic lineages of the same game-farm mallard breed (Lavretsky et al. 2023). Nevertheless, the peak population sizes and times associated with changing N_E differed. Nearly identical demographic histories were recovered for domestic genomes from North America (Figure 4.2A) and Hawaii (Figure 4.4A), with N_E exponentially increasing up to ~700,000 years before present (YBP) before gradually declining to a contemporary effective population size of 10,000 individuals today. For the domestic genome from New Zealand, I recovered an increasing N_E to around 270,000 YBP before gradually decreasing to a contemporary effective population size of 230,000 today.

Finally, genomes representing the various domestic x wild hybrids tended to recover demographic histories closely tied to the parental taxa that is overwhelmingly represented in its genome. Although hybrid genomes did tend to show intermediate estimates in N_E across time relative to their parental taxa, both N_E and time were substantially distorted (Figures 4.2-4.4). For example, genomes comprised of more wild mallard in North America (i.e., Hybrid 1 & 2;

Figure 4.2) showed a similar, though depressed pattern in N_E, but with these patterns consistently pushed forward in time. Interestingly, Hybrid 3 that was closest to 50:50 ancestry split between game-farm and wild mallard based on partial genome data (Table 4.1; Lavretsky et al. 2023) showed the expected combination effect by essentially depressing the wild signature through time but without a time-shift in those patterns (Figure 4.2). Conversely, genomes from Hawaii simply showed depressed estimates of N_E but all patterns pushed forward in time (Figure 4.4). Finally, hybrid genomes from New Zealand generally followed the exponential increase in N_E through time as seen in the New Zealand grey duck (Figure 4.3).

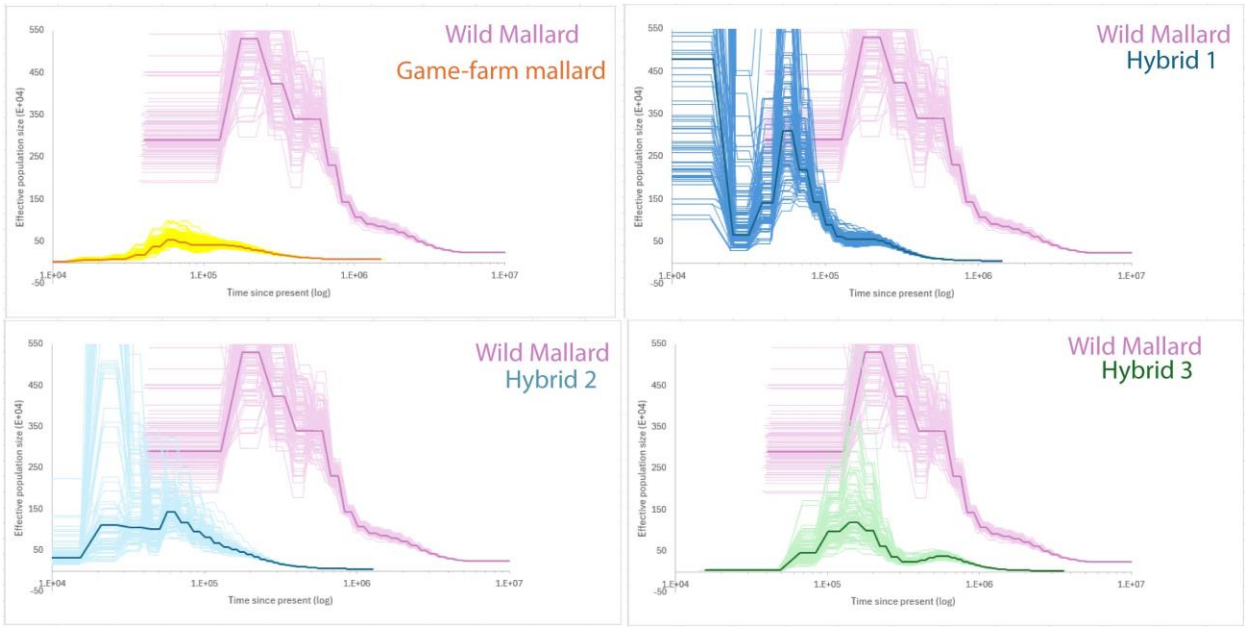

Figure 4.2: PSMC results from wild and domestic mallards and their hybrids from North America. In each analysis, the dark denotes the average, while the lighter lines indicate the bootstraps.

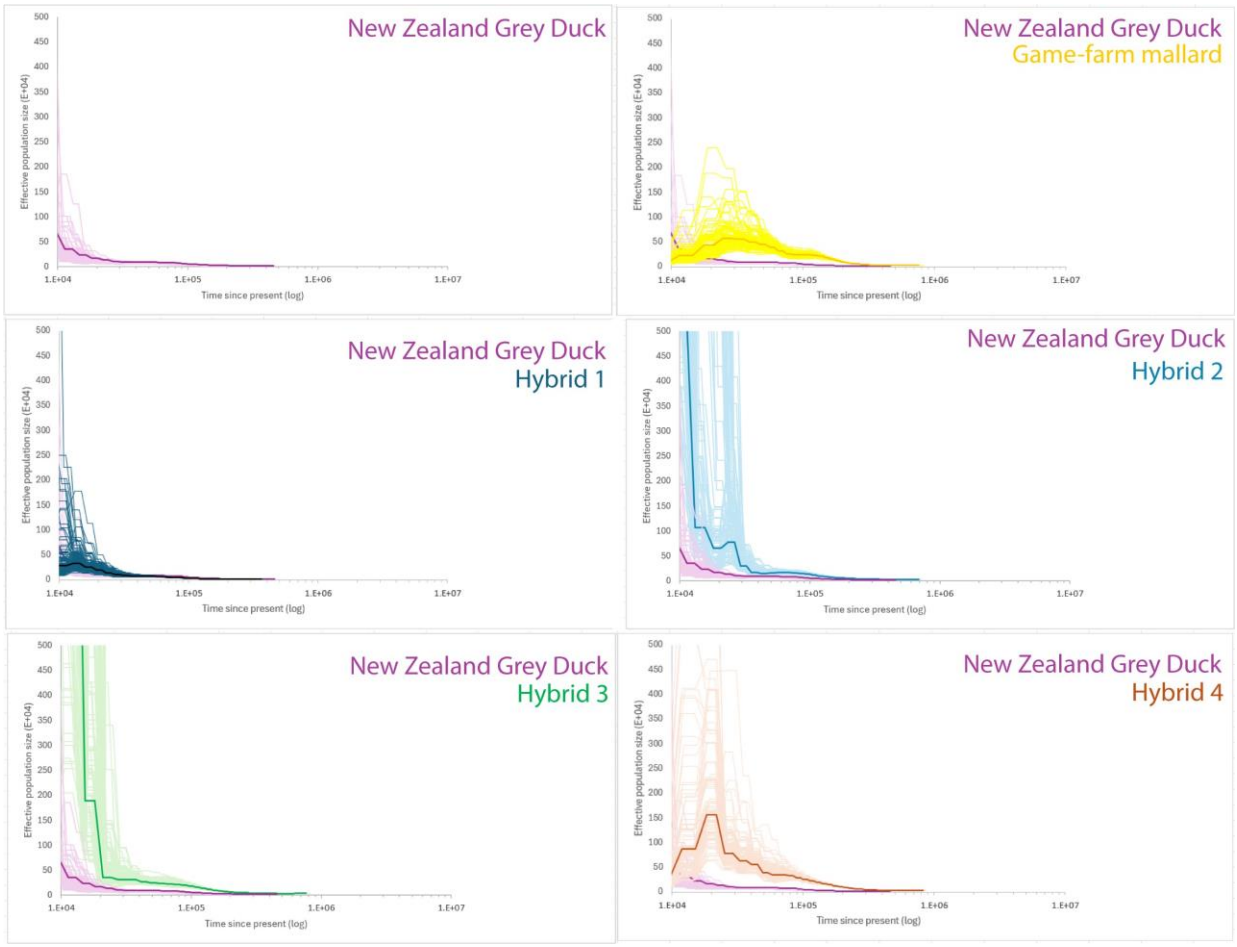

Figure 4.3: PSMC results from New Zealand Grey Duck, domestic mallard, and their hybrids from New Zealand. In each analysis, the dark denotes the average, while the lighter lines indicate the bootstraps.

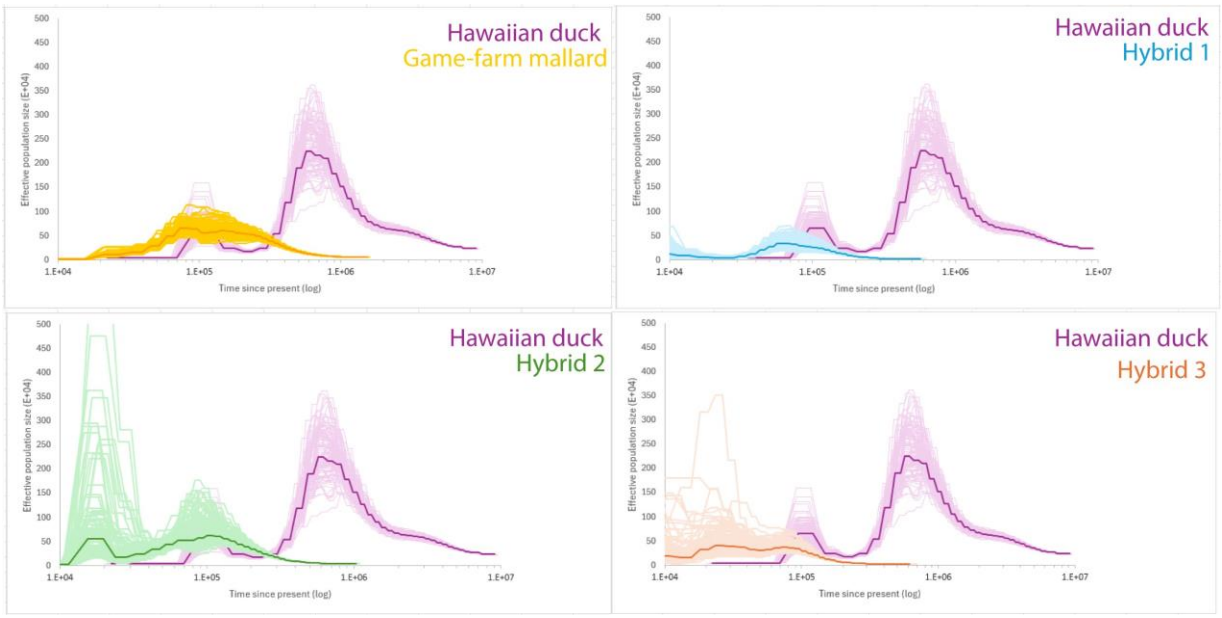

Figure 4.4: PSMC results from Hawaiian duck, domestic mallard, and their hybrids from Hawaii. In each analysis, the dark denotes the average, while the lighter lines indicate the bootstraps.

DISCUSSION

The history of supplementing the game-farm mallard breed across the world has indeed resulted in wide-spread anthropogenic hybridization, and I conclude to have negatively impacted the diversity of their wild congener's genomes. Generally, artificial selection for traits either specifically targeted by people or unknowingly pressured in a domestic setting often results in significant changes in population traits and their associated genetic variation (Andersson & Georges, 2004; X. Li et al., 2020; Söderquist et al., 2014). In addition to strong artificial selection, domestication and breed selection innately result in sequential bottlenecking; both resulting in the increasing chance of inheriting identical-by-descent genetic material (Curik et al., 2014). Together, these pressures tend to exacerbate genetic diversity loss, increasing inbreeding depression, and resulting in extended homozygous regions in the offspring's genome (i.e., increasing ROH) (Bosse et al., 2019; Curik et al., 2014; Willoughby et al., 2015). Indeed, I

recovered expected summary statistics and depressed effective population size estimates for the genome representing the game-farm mallard that is currently being bred for release in North America and Eurasia (Table 4.1; Figs. 4.1-4.2). In fact, these genomic changes have recently been associated with morphological traits in which game-farm mallards tended to possess shorter, wider, and taller bills than their wild counterparts, which could influence their foraging behavior and diet preferences in the wild (Söderquist et al., 2014, Halligan 2024). Such morphological and behavioral traits are likely only some of the traits that were born in the domestic setting, but regardless, the movement of these traits that are likely maladaptive in the wild may at least partially explain the ever-declining wild mallard populations of North America where game-farm mallard releases are highest (Heusmann, 1991, USFW 2013). I expect any traits with maladaptive consequences in the wild to be under strong purifying selection (Gering et al., 2019); however, the continued release of game-farm mallards evidently perpetuates significant levels of gene flow that may counteract selective forces (Lavretsky, Mohl, et al., 2023; Söderquist et al., 2017). Importantly, with only an average of ~2% of samples considered wild, the eastern portion of North America is now considered a game-farm x wild mallard hybrid swarm (Lavretsky et al. 2023). And thus, without a sufficiently large enough wild gene pool, the resulting offspring are expected to continue to have a combination of hybrid ancestry that may further prevent the loss of those maladaptive traits across generations (Schummer et al., 2023). Additionally, it is also possible that some traits born in the domestic setting may be neutral or even positively selected for in more urban settings (Driscoll et al., 2009), and thus, perpetuated among growing populations of urban waterfowl. Future work will benefit from extensive genomic sampling of individuals representing the spectrum of game-farm x wild mallard hybrids found in North America to test (1) whether the diversity statistics remain correlated (Fig. 4.1),

but more importantly, (2) to test the strength of selection by determining the speed (i.e., generations) in which domestically-derived genetic variation is lost in the wild and whether there is heterogeneity in selective pressure across urban and rural habitats. Regardless, I suggest that the domestication process negatively affected the diversity of the game-farm mallard's genome, and their release and subsequent interbreeding with wild congeners might result in offspring with genomes that are generally more inbred than their wild parent (Table 4.1; Fig. 4.1), and which has serious conservation implications.

Unlike North America where I find expected worsening genetic diversity values among genomes with increasing domestic ancestry, genomes from Hawaii and New Zealand provided uniquely different patterns (Table 4.1; Fig. 4.1). I note that while game-farm mallards continue to be released in North America, major releases have largely ceased and most mallards in Hawaii and New Zealand comprise respective self-sustaining feral populations today (Engilis Jr. et al., 2020; Lavretsky, Mohl, et al., 2023; USGS, 2007; Williams, 2017). The differing history of supplementation likely explains the somewhat unexpected genetic patterns for these Island genomes in which I either recover reverse correlations or no correlations between summary statistics and levels of domestic ancestry in New Zealand and Hawaii, respectively (Table 4.1; Fig. 4.1). However, the expected pattern persists in runs of homozygosity, where greater domestic ancestry leads to longer runs; albeit with a lower correlation in New Zealand and no significant statistical effect in Hawaii. I posit that the lack of correlation in Hawaii is likely a result of genetic drift being the overwhelming force because of feral populations being maintained in relatively small sizes and evidently absent of gene flow between locations (Wells et al. 2019); both of which explain the generally low levels of genetic diversity, large ROH, and high inbreeding coefficients regardless of the proportional contribution of the domestic parent

(Table 4.1; Fig. 4.1). Conversely, I find that the genome representing New Zealand's feral mallard population harbored higher nucleotide diversity and lower inbreeding coefficients than endemic grey ducks (Table 4.1; Fig. 4.1). Again, whereas ROH did increase with the amount of domestic ancestry among New Zealand samples, these runs were comparable to those recovered in wild mallards (Table 4.1). Together, I hypothesize that a combination of successive recombination, as well as likely adaptive introgression from New Zealand grey ducks and local selective pressures have already refined the genome of New Zealand's feral mallard population, explaining why genetically they no longer reflect their source population (Lavretsky, Mohl, et al., 2023), their exponentially increasing and self-sustaining populations (Brown, 2021; Caithness et al., 1991; Williams, 2017), and the generally positive outcomes on genomic diversity found here. In the end, results from Hawaii and particularly New Zealand, demonstrate that the negative genomic outcomes from initial hybridization between domestic and wild congeners as being observed in North America might be refined with sufficient generations and a lack or decline in continued gene flow of the source domestic lineage. However, I also acknowledge Islands that generally have less complex ecologies may provide a backdrop for feralization that is absent in landscapes like North America. Consequently, understanding how feral and/or hybrid genomes may evolve in the changing ecologies of the Anthropocene cannot be considered singular, requiring the consideration of the ecological backdrop in which they are doing so (Ottenburghs, 2021).

Domestic and hybrid genomes bias demographic inferences

Reconstructing demographic histories and assessing contemporary levels of effective population size from genomic data has become integral in evolutionary and conservation genetic

studies (Hu et al. 2020, Saremi et al. 2019; Wang et al, 2021). Not only are such analyses helpful to understand whether current standing genetic variation is a result of evolutionary or more contemporary trends, but establishing today's effective population size is critical and often an important consideration when determining conservation priorities (Brüniche-Olsen et al., 2021; Gabrielli et al., 2024). Given that PSMC analyses that provide such information are done so using a single genome, what that genome represents is equally important (Hung et al., 2014; Nadachowska-Brzyska et al., 2015; Zhao et al., 2013). Towards this, I provide evidence that genomes subjugated to the domestication process or of hybrid origins substantially bias demographic inferences (Figs. 4.2-4.4; (Brown et al., 2022; Lu et al., 2022; Nikolic et al., 2020). First, my results confirm that the domestication process that results in ever-increasingly inbred lines (Table 4.1) causes demographic histories to be generally depressed, with similar patterns of slightly increasing N_E in deep time followed by decreasing trends in N_E to more recent times (Figs. 4.2-4.4). Next, whereas I recovered expected trends for genomes of hybrid individuals that followed demographics of their wild counterparts but with depressed values of N_E, I unexpectedly also found that these trends were shifted forward in time. Thus, not only do inbred or hybrid genomes pose a serious consequence in estimating contemporary levels of N_E, but that any inferences about divergence time between lineages is equally biased (H. Li & Durbin, 2011; MacLeod et al., 2013; C. A. Martin et al., 2023). This is especially concerning for genomes representing species but constructed from individuals born in captive settings (e.g., zoos) where both unintentional selective pressures, potential bottlenecking, and even admixture may cause such genomes to significantly diverge from the original wild source. Together, researchers should always consider the individual's source prior to making effort to build genomes and

reconstruct demographic histories, especially if such information is to be used to infer evolutionary histories or conservation planning.

REFERENCES

Adavoudi, R., & Pilot, M. (2022). Consequences of Hybridization in Mammals: A Systematic Review. *Genes*, *13*(1), Article 1. https://doi.org/10.3390/genes13010050

Alerstam, T., & Högstedt, G. (1982). Bird Migration and Reproduction in Relation to Habitats for Survival and Breeding. *Ornis Scandinavica (Scandinavian Journal of Ornithology)*, *13*(1), 25–37. https://doi.org/10.2307/3675970

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: Setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Andersson, L., & Georges, M. (2004). Domestic-animal genomics: Deciphering the genetics of complex traits. *Nature Reviews. Genetics*, *5*(3), 202–212. https://doi.org/10.1038/nrg1294

Arnold, T. W., & Clark, R. G. (1996). Survival and Philopatry of Female Dabbling Ducks in Southcentral Saskatchewan. *The Journal of Wildlife Management*, *60*(3), 560–568. https://doi.org/10.2307/3802073

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bosse, M., Megens, H.-J., Derks, M. F. L., de Cara, Á. M. R., & Groenen, M. A. M. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evolutionary Applications*, *12*(1), 6–17. https://doi.org/10.1111/eva.12691

Brown, J. I. (2021). *Landscape Population And Evolutionary Genomics Of Several Closely Related Species Of Mallard-Like Ducks*.

Brown, J. I., Harrigan, R. J., & Lavretsky, P. (2022). Evolutionary and ecological drivers of local adaptation and speciation in a North American avian species complex. *Molecular Ecology*, *31*(9), 2578–2593. https://doi.org/10.1111/mec.16423

Brüniche-Olsen, A., Kellner, K. F., Belant, J. L., & DeWoody, J. A. (2021). Life-history traits and habitat availability shape genomic diversity in birds: Implications for conservation. *Proceedings. Biological Sciences*, *288*(1961), 20211441. https://doi.org/10.1098/rspb.2021.1441

Buckley, Y. M., & Catford, J. (2016). Does the biogeographic origin of species matter? Ecological effects of native and non-native species and the use of origin to guide management. *Journal of Ecology*, *104*(1), 4–17. https://doi.org/10.1111/1365-2745.12501

Caithness, T., Williams, M., & Nichols, J. D. (1991). Survival and Band Recovery Rates of Sympatric Grey Ducks and Mallards in New Zealand. *The Journal of Wildlife Management*, *55*(1), 111–118. https://doi.org/10.2307/3809247

Casas, F., Mougeot, F., Sánchez-Barbudo, I., Dávila, J. A., & Viñuela, J. (2012). Fitness consequences of anthropogenic hybridization in wild red-legged partridge (Alectoris rufa, Phasianidae) populations. *Biological Invasions*, *14*(2), 295–305. https://doi.org/10.1007/s10530-011-0062-3

Champagnon, J., Crochet, P.-A., Kreisinger, J., Čížková, D., Gauthier-Clerc, M., Massez, G., Söderquist, P., Albrecht, T., & Guillemain, M. (2013). Assessing the genetic impact of massive restocking on wild mallard. *Animal Conservation*, *16*(3), 295–305. https://doi.org/10.1111/j.1469-1795.2012.00600.x

Crispo, E., Moore, J.-S., Lee-Yaw, J. A., Gray, S. M., & Haller, B. C. (2011). Broken barriers: Human-induced changes to gene flow and introgression in animals: an examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *33*(7), 508–518. https://doi.org/10.1002/bies.201000154

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Drilling, N., Titman, R. D., & McKinney, F. (2020). Mallard (Anas platyrhynchos), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.mallar3.01species_shared.bow.project_name

Driscoll, C. A., Macdonald, D. W., & O'Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proceedings of the National Academy of Sciences of the United States of America*, *106 Suppl 1*(Suppl 1), 9971–9978. https://doi.org/10.1073/pnas.0901586106

Engilis Jr., A., Uyehara, K. J., & Giffin, J. G. (2020). Hawaiian Duck (Anas wyvilliana), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.hawduc.01species_shared.bow.project_name

Gabrielli, M., Leroy, T., Salmona, J., Nabholz, B., Milá, B., & Thébaud, C. (2024). Demographic responses of oceanic island birds to local and regional ecological disruptions revealed by whole-genome sequencing. *Molecular Ecology*, *33*(4), e17243. https://doi.org/10.1111/mec.17243

Gering, E., Incorvaia, D., Henriksen, R., Wright, D., & Getty, T. (2019). Maladaptation in feral and domesticated animals. *Evolutionary Applications*, *12*(7), 1274–1286. https://doi.org/10.1111/eva.12784

Gillespie, G. D. (1985). Hybridization, Introgression, and Morphometric Differentiation between Mallard (Anas platyrhynchos) and Grey Duck (Anas superciliosa) in Otago, New Zealand. *The Auk*, *102*(3), 459–469. https://doi.org/10.1093/auk/102.3.459

Grabenstein, K. C., Otter, K. A., Burg, T. M., & Taylor, S. A. (2023). Hybridization between closely related songbirds is related to human habitat disturbance. *Global Change Biology*, *29*(4), 955–968. https://doi.org/10.1111/gcb.16476

Grabenstein, K. C., & Taylor, S. A. (2018). Breaking Barriers: Causes, Consequences, and Experimental Utility of Human-Mediated Hybridization. *Trends in Ecology & Evolution*, *33*(3), 198–212. https://doi.org/10.1016/j.tree.2017.12.008

Guo, Q. (2014). Plant hybridization: The role of human disturbance and biological invasion. *Diversity and Distributions*, *20*(11), 1345–1354. https://doi.org/10.1111/ddi.12245

Hedrick, P. W. (2015). Heterozygote Advantage: The Effect of Artificial Selection in Livestock and Pets. *Journal of Heredity*, *106*(2), 141–154. https://doi.org/10.1093/jhered/esu070

Heusmann, H. W. (1991). *The History and Status of the Mallard in the Atlantic Flyway*. *19*(1), 14–22.

Hung, C.-M., Shaner, P.-J. L., Zink, R. M., Liu, W.-C., Chu, T.-C., Huang, W.-S., & Li, S.-H. (2014). Drastic population fluctuations explain the rapid extinction of the passenger pigeon. *Proceedings of the National Academy of Sciences*, *111*(29), 10636–10641. https://doi.org/10.1073/pnas.1401526111

Kidd, A. G., Bowman, J., Lesbarrères, D., & Schulte-Hostedde, A. I. (2009). Hybridization between escaped domestic and wild American mink (Neovison vison). *Molecular Ecology*, *18*(6), 1175–1186. https://doi.org/10.1111/j.1365-294X.2009.04100.x

Lamont, B. B., He, T., Enright, N. J., Krauss, S. L., & Miller, B. P. (2003). Anthropogenic disturbance promotes hybridization between Banksia species by altering their biology. *Journal of Evolutionary Biology*, *16*(4), 551–557. https://doi.org/10.1046/j.1420-9101.2003.00548.x

Lavretsky, P., Hernández, F., Swale, T., & Mohl, J. E. (2023). Chromosomal-level reference genome of a wild North American mallard (Anas platyrhynchos). *G3 Genes|Genomes|Genetics*, *13*(10), jkad171. https://doi.org/10.1093/g3journal/jkad171

Lavretsky, P., McInerney, N. R., Mohl, J. E., Brown, J. I., James, H. F., McCracken, K. G., & Fleischer, R. C. (2020). Assessing changes in genomic divergence following a century of human-mediated secondary contact among wild and captive-bred ducks. *Molecular Ecology*, *29*(3), 578–595. https://doi.org/10.1111/mec.15343

Lavretsky, P., Mohl, J. E., Söderquist, P., Kraus, R. H. S., Schummer, M. L., & Brown, J. I. (2023). The meaning of wild: Genetic and adaptive consequences from large-scale releases of domestic mallards. *Communications Biology*, *6*(1), 1–15. https://doi.org/10.1038/s42003-023-05170-w

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), Article 7357. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, X., Yang, J., Shen, M., Xie, X.-L., Liu, G.-J., Xu, Y.-X., Lv, F.-H., Yang, H., Yang, Y.-L., Liu, C.-B., Zhou, P., Wan, P.-C., Zhang, Y.-S., Gao, L., Yang, J.-Q., Pi, W.-H., Ren, Y.-L., Shen, Z.-Q., Wang, F., … Li, M.-H. (2020). Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nature Communications*, *11*(1), 2815. https://doi.org/10.1038/s41467-020-16485-1

Lu, C.-W., Yao, C.-T., & Hung, C.-M. (2022). Domestication obscures genomic estimates of population history. *Molecular Ecology*, *31*(3), 752–766. https://doi.org/10.1111/mec.16277

MacLeod, I. M., Larkin, D. M., Lewin, H. A., Hayes, B. J., & Goddard, M. E. (2013). Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. *Molecular Biology and Evolution*, *30*(9), 2209–2223. https://doi.org/10.1093/molbev/mst125

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and

contemporary inbreeding across diverging populations of an island-colonizing bird. *Molecular Ecology*, *32*(8), 1972–1989. https://doi.org/10.1111/mec.16865

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

McFarlane, S. E., & Pemberton, J. M. (2019). Detecting the True Extent of Introgression during Anthropogenic Hybridization. *Trends in Ecology & Evolution*, *34*(4), 315–326. https://doi.org/10.1016/j.tree.2018.12.013

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. *Current Biology: CB*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044

Nikolic, N., Liu, S., Jacobsen, M. W., Jónsson, B., Bernatchez, L., Gagnaire, P.-A., & Hansen, M. M. (2020). Speciation history of European (Anguilla anguilla) and American eel (A. rostrata), analysed using genomic data. *Molecular Ecology*, *29*(3), 565–577. https://doi.org/10.1111/mec.15342

Ottenburghs, J. (2021). The genic view of hybridization in the Anthropocene. *Evolutionary Applications*, *14*(10), 2342–2360. https://doi.org/10.1111/eva.13223

Randi, E. (2008). Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology, 17*(1), 285–293. https://doi.org/10.1111/j.1365-294X.2007.03417.x

Rhymer, J. M., & Simberloff, D. (1996). Extinction by hybridization and introgression. *Annual Review of Ecology, Evolution, and Systematics, 27*(Volume 27, 1996), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Salvatori, V., Godinho, R., Braschi, C., Boitani, L., & Ciucci, P. (2019). High levels of recent wolf × dog introgressive hybridization in agricultural landscapes of central Italy. *European Journal of Wildlife Research, 65*(5), 73. https://doi.org/10.1007/s10344-019-1313-3

Schummer, M. L., Simpson, J., Shirkey, B., Kucia, S. R., Lavretsky, P., & Tozer, D. C. (2023). Population genetics and geographic origins of mallards harvested in northwestern Ohio. *PLOS ONE*, *18*(3), e0282874. https://doi.org/10.1371/journal.pone.0282874

Smith, G. W., & Reynolds, R. E. (1992). Hunting and Mallard Survival, 1979-88. *The Journal of Wildlife Management*, *56*(2), 306–316. https://doi.org/10.2307/3808827

Söderquist, P., Elmberg, J., Gunnarsson, G., Thulin, C.-G., Champagnon, J., Guillemain, M., Kreisinger, J., Prins, H. H. T., Crooijmans, R. P. M. A., & Kraus, R. H. S. (2017). Admixture between released and wild game birds: A changing genetic landscape in European mallards (Anas platyrhynchos). *European Journal of Wildlife Research*, *63*(6), 98. https://doi.org/10.1007/s10344-017-1156-8

Söderquist, P., Norrström, J., Elmberg, J., Guillemain, M., & Gunnarsson, G. (2014). Wild mallards have more “goose-like” bills than their ancestors: A case of anthropogenic influence? *PloS One, 9*(12), e115143. https://doi.org/10.1371/journal.pone.0115143

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., Heredia, S. M., Hahn, M. A., Caseys, C., Bock, D. G., & Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

USGS. (2007). *Hawaiian Duck's future threatened by feral mallards* (Fact Sheet) [Fact Sheet].

Williams, M. (2017). *The changing relative abundance of grey duck (Anas superciliosa) and mallard (A. platyrhynchos) in New Zealand*. *64*, 211–228.

Willoughby, J. R., Fernandez, N. B., Lamb, M. C., Ivy, J. A., Lacy, R. C., & DeWoody, J. A. (2015). The impacts of inbreeding, drift and selection on genetic diversity in captive breeding populations. *Molecular Ecology*, *24*(1), 98–110. https://doi.org/10.1111/mec.13020

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., Hu, Y., He, W., Zhang, S., Fan, W., Zhu, L., Li, D., Zhang, X., Chen, Q., Zhang, H., Zhang, Z., Jin, X., Zhang, J., Yang, H., … Wei, F. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, *45*(1), 67–71. https://doi.org/10.1038/ng.2494

Appendix

SUPPLEMENTAL TABLES

Table S1.1. Sample information. Museums: UTEP: University of Texas at El Paso, UCD: University of California, Davis

Sample Size	Museum ID	Species	Location	State/ Province	Country	Latitude	Longitude	Sex
1	UTEP-Bird 3056	Wild Mallard	Sierra Caballo Reservoir	New Mexico	US	32.953968	-107.295992	M
1	UCD - 25385-2	Laysan Duck	Sand Island	Hawaii	US	28.2097	-177.3754	M
1	UCD - 25386-4	Hawaiian Duck	Kauai Island	Hawaii	US	22.03616	-159.77894	M

Table 1.2. Sequencing and assembly results of my Laysan duck and Hawaiian duck reference genomes.

Stats	Laysan Duck	Hawaiian Duck
Shotgun library- Input Assembly		
Genome Coverage	34.06x	71.61x
Total Length	1,042.16 Mb	1,042.16 Mb
Longest scaffold	1,060,896 bp	1,060,896 bp
# of scaffolds	36,568	36,568
# of scaffolds >1kb	36,568	36,568
N50	0.065 Mb	0.065 Mb
L50	4,254 scaffolds	4,254 scaffolds
N90	0.015 Mb	0.015 Mb
L90	17,147 scaffolds	17,147 scaffolds
# of gaps	19,503	19,503
Percent of genome in gaps	0.24%	0.24%
Single copy BUSCO	70.3%	70.3%
Duplicated BUSCO	0.7%	0.7%
Chicago library -Dovetail HiRise Assembly		
Genome Coverage	99.9	99.9
Total Length	1,044.86 Mb	1,045.44 Mb
Longest scaffold	7,869,839 bp	59,612,633 bp
# of scaffolds	9,721	3,859
# of scaffolds >1kb	9,715	3,858
GC %	41.01	41.01
N50	0.761 Mb	8.035 Mb
L50	361 scaffolds	38 scaffolds

N90	0.116 Mb	1.990 Mb
L90	1,681 scaffolds	136 scaffolds
# of gaps	46,478	52,224
Percent of genome in gaps	0.49%	0.55%
Single copy BUSCO	80.5%	81.5%
Duplicated BUSCO	0.7%	0.7%

Table S1.3. Alignment report of the Laysan duck, Hawaiian duck, and mallard nuclear autosomal, Z-chromosome and mitochondrial genomes using BWA. Coverage statistics were calculated using SAMtools coverage.

Stats	Laysan duck	Hawaiian duck	Wild Mallard
	Autosomes		
No. reads	23258274	22332019	20777591
Covered bases (Mbp)	33.96	34.02	34.07
Alignment (%)	98.91	99.10	99.31
Mean coverage (x)	83.9	84.6	67.4
	Z- chromosome		
No. reads	58033671	55971181	54179853
Covered bases (Mbp)	78.51	78.63	78.72
Alignment (%)	99.13	99.28	99.39
Mean coverage (x)	90.7	85.7	78.1
	Mitochondrial DNA		
No. reads	8993147	13134573	13064630
Covered bases (kbp)	16.6	16.6	16.6
Alignment (%)	100	100	100
Mean coverage (x)	6.86E+04	1.00E+05	9.85E+04

Table S1.4. Number of homozygous sites (n_hom) in the Hawaiian duck and its parent species. Number of heterozygous sites (n_het) and probability of heterozygosity (p_het) in the Hawaiian duck.

	Autosomes				**Z-chromosome**			
species	**n_hom (p1)**	**n_het**	**n_hom (p2)**	**p_het**	**n_hom (p1)**	**n_het**	**n_hom (p2)**	**p_het**
Laysan duck	624665	0	0	0	44007	0	0	0
Hawaiian duck	187007	217731	218230	0.35	14773	18918	10252	0.431

Mallard	0	0	624665	0	0	0	44007	0

Table S1.5. Results from Patterson's D test for introgression between lineages with block Jack-knifed SE estimates and Z scores per chromosome. Z scores are highlighted in bold to indicate significance of D' statistics (Z score >4).

	P1	**P2**	**P3**	**P1**	**P2**	**P3**
	Laysan	Hawaiian	Mallard	Mallard	Hawaiian	Laysan
Chromo-some	**Patterson's D**	**Jack-knifed SE**	**Z score**	**Patterson's D**	**Jack-knifed SE**	**Z score**
1	0.1100	0.0096	**11.4840**	0.47719	0.01942	**24.58534**
2	0.1077	0.0097	**11.1508**	0.531105	0.021129	**25.14848**
3	0.1472	0.0141	**10.4674**	0.512613	0.023152	**22.15291**
4	0.0858	0.0144	**5.9865**	0.455511	0.030813	**14.79431**
5	0.1293	0.0164	**7.8789**	0.422436	0.034448	**12.27945**
6	0.1892	0.0345	**5.4655**	0.431797	0.040209	**10.82099**
7	0.1034	0.0150	**6.9265**	0.617972	0.035555	**17.39015**
8	0.1424	0.0207	**6.8687**	0.520456	0.040209	**10.82099**
9	0.1107	0.0200	**5.5225**	0.537584	0.04244	**12.6934**
10	0.0446	0.0278	1.5954	0.537823	0.061265	**8.747896**
11	0.1269	0.0231	**5.5776**	0.565942	0.03832	**14.78225**
12	0.1152	0.0247	**4.7417**	0.677068	0.045321	**14.97519**
13	0.1366	0.0179	**7.6366**	0.450005	0.043848	**10.28427**
14	0.1307	0.0164	**8.0271**	0.582923	0.044551	**13.12493**
15	0.1522	0.0261	**5.8175**	0.567523	0.057959	**9.845429**
16	0.1752	0.0251	**6.9419**	0.606821	0.036673	**16.51777**
17	0.0611	NA	NaN	-0.00294	NA	NaN
18	0.1286	0.0234	**5.5801**	0.599886	0.061281	**9.882182**
19	0.0748	0.0181	**4.1348**	0.63986	0.026945	**23.73743**
20	0.1296	0.0266	**4.8598**	0.577506	0.042664	**13.5519**
21	0.1169	0.0291	**4.0811**	0.527957	0.038939	**13.53409**
22	0.1262	0.0477	2.7059	0.695948	0.043665	**15.9443**
23	0.1293	0.0197	**6.7220**	0.515469	0.058129	**8.844522**
24	0.1188	0.0221	**5.4617**	0.355219	0.096766	3.753058
25	0.1529	0.0299	**5.1458**	0.472717	0.08828	**5.427744**
26	0.1197	0.0161	**7.6823**	0.370227	0.192362	2.126373
27	0.1496	0.0231	**6.5219**	0.519696	0.067239	**7.728536**
28	0.0993	0.0220	**4.4964**	0.549912	0.051187	**10.81234**
29	0.1454	0.0247	**6.0142**	0.611373	0.093957	**6.500505**

Table S1.6. List of genes and associated GO Terms in outlier windows for LH, MH, and PH.

Windows Type	Chr	Start	End	Gene	Description	GO Term
LH	chr1	31100001	31200000	g674.t1	hypothetical protein	GO:0003964,GO:0006278
LH	chr1	31100001	31200000	g675.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like, partial	
LH	chr1	31100001	31200000	g676.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016772,GO:0044237
LH	chr1	31100001	31200000	g677.t1	-	
LH	chr1	33400001	33500000	g746.t1	hypothetical protein	GO:0000276,GO:0015078,GO:0006811, GO:0015986
LH	chr1	33400001	33500000	g747.t1	methylase	
LH	chr1	33400001	33500000	g748.t1	reverse hypothetical protein	GO:0003964,GO:0006278
LH	chr1	34600001	34700000	g790.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0003964,GO:0006278
LH	chr1	34600001	34700000	g791.t1	hypothetical protein llap_18751	GO:0003964,GO:0016301,GO:0006278, GO:0016310
LH	chr1	34600001	34700000	g792.t1	PREDICTED: DEP domain-containing protein 1B, partial	
LH	chr1	34600001	34700000	g793.t1	hypothetical protein llap_1060	
LH	chr1	34700001	34800000	g794.t1	hypothetical protein, partial	
LH	chr1	34700001	34800000	g795.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
LH	chr1	34700001	34800000	g796.t1	putative inactive tRNA-specific adenosine deaminase-like protein 3	
LH	chr1	34700001	34800000	g797.t1	hypothetical protein C5E46_35750	
LH	chr1	34700001	34800000	g798.t1	reverse hypothetical protein	GO:0003964,GO:0006278

LH	chr1	34700001	34800000	g799.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
LH	chr1	34900001	35000000	g804.t1	hypothetical protein llap_3840	GO:0016772,GO:0044237
LH	chr1	34900001	35000000	g805.t1	vps10 domain-containing receptor 2	
LH	chr1	34900001	35000000	g806.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
LH	chr1	34900001	35000000	g807.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0003964,GO:0006278
LH	chr1	34900001	35000000	g808.t1	hypothetical protein llap_9744	
LH	chr1	35300001	35400000	g818.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
LH	chr1	35300001	35400000	g819.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
LH	chr1	35300001	35400000	g820.t1	pol- hypothetical protein	GO:0016021,GO:0003964,GO:0006278
LH	chr1	35300001	35400000	g821.t1	hypothetical protein DUI87_05782	GO:0003824,GO:0030234
LH	chr1	35300001	35400000	g822.t1	rna-directed dna polymerase from mobile element jockey-like	
LH	chr1	35300001	35400000	g823.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016020,GO:0003964,GO:0006278, GO:0007165
LH	chr1	35300001	35400000	g824.t1	hypothetical protein N327_05121, partial	
LH	chr1	35300001	35400000	g825.t1	PREDICTED: enteropeptidase	
LH	chr1	35300001	35400000	g826.t1	hypothetical protein	GO:0003964,GO:0006278
LH	chr1	35400001	35500000	g828.t1	reverse transcriptase family protein, partial	GO:0003964,GO:0006278
LH	chr1	35400001	35500000	g829.t1	hypothetical protein	
LH	chr1	35400001	35500000	g830.t1	hypothetical protein	
LH	chr1	35400001	35500000	g831.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003964,GO:0006278
LH	chr1	35400001	35500000	g832.t1	hypothetical protein C5E44_35305, partial	
LH	chr1	35400001	35500000	g833.t1	PREDICTED: enteropeptidase	
LH	chr1	37800001	37900000	g924.t1	rna-directed dna polymerase from mobile	GO:0003824

					element jockey-like	
LH	chr1	37800001	37900000	g925.t1	reverse transcriptase family protein, partial	GO:0003964,GO:0006278
LH	chr1	37800001	37900000	g926.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like	GO:0030234
LH	chr1	38500001	38600000	g958.t1	hypothetical protein AC249_AIPGENE27315	
LH	chr1	38500001	38600000	g959.t1	adaptin ear-binding coat-associated protein 1	GO:0003824,GO:0004867,GO:0010951
LH	chr1	38500001	38600000	g960.t1	hypothetical protein DUI87_27591	GO:0004190,GO:0006508
LH	chr1	85000001	85100000	g2139.t1	Signal recognition particle receptor subunit alpha, partial	
LH	chr1	85000001	85100000	g2140.t1	PREDICTED: protein FAM118B isoform X1	
LH	chr1	85000001	85100000	g2141.t1	Toll/interleukin-1 receptor domain-containing adapter protein, partial	GO:0003824
LH	chr1	1.28E+08	1.28E+08	g3031.t1	MANSC domain-containing protein 1	GO:0016021
LH	chr1	1.28E+08	1.28E+08	g3033.t1	dual specificity protein phosphatase 16	GO:0005654,GO:0005737,GO:0008432, GO:0017017,GO:0017018,GO:0048273, GO:0006470,GO:0043508,GO:0045204
LH	chr1	1.49E+08	1.49E+08	g3528.t1	rna-directed dna polymerase from mobile element jockey-like	
LH	chr1	1.49E+08	1.49E+08	g3529.t1	rna-directed dna polymerase from mobile element jockey-like	
LH	chr1	1.49E+08	1.49E+08	g3530.t1	PREDICTED: uncharacterized protein LOC106045565	
LH	chr1	1.49E+08	1.49E+08	g3531.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
LH	chr1	1.49E+08	1.49E+08	g3532.t1	CUB and sushi domain-containing protein 1	GO:0008046,GO:0016199,GO:0035385
LH	chr1	1.49E+08	1.49E+08	g3533.t1	PREDICTED: enteropeptidase	
LH	chr1	1.49E+08	1.49E+08	g3534.t1	hypothetical protein A6L34_02820	
LH	chr1	1.61E+08	1.61E+08	g3792.t1	hypothetical protein HGM15179_008129	GO:0005634,GO:0005737,GO:0000049, GO:0004672,GO:0005524,GO:0031267,

						GO:0006468,GO:0006886,GO:0071528
LH	chr1	1.61E+08	1.61E+08	g3793.t1	60S ribosomal protein L18a	GO:0022625,GO:0042788,GO:0003735, GO:0002181
LH	chr1	1.61E+08	1.61E+08	g3794.t1	KICSTOR complex protein C12orf66 homolog isoform X1	GO:0005764,GO:0045171,GO:0140007, GO:0034198,GO:0042149,GO:0061462, GO:1904262
LH	chr2	42600001	42700000	g5303.t1	protein-L-isoaspartate O-methyltransferase domain-containing protein 1	GO:0005737,GO:0004719,GO:0006479
LH	chr2	42600001	42700000	g5304.t1	hypothetical protein C5E44_35255	GO:0003964,GO:0006278
LH	chr2	51100001	51200000	g5456.t1	hypothetical protein DUI87_07715	GO:0016579,GO:0070628,GO:0007268, GO:0000502,GO:0007266,GO:0045202, GO:0004674,GO:0006468,GO:0005886, GO:0009986,GO:0004843,GO:0005524, GO:0005856,GO:0004866,GO:0005829, GO:0031267,GO:0031410,GO:0010951, GO:0050920,GO:1903070
LH	chr2	94200001	94300000	g6346.t1	M-phase-specific PLK1-interacting protein	
LH	chr2	94700001	94800000	g6351.t1	hypothetical protein DUI87_22404	GO:0005615,GO:0003964,GO:0004867, GO:0006278,GO:0010951
LH	chr2	95200001	95300000	g6356.t1	PREDICTED: paired box protein Pax-6 isoform X4	
LH	chr2	95300001	95400000	g6357.t1	hypothetical protein llap_4974	
LH	chr2	95300001	95400000	g6358.t1	hypothetical protein Y1Q_0017503	GO:0005634,GO:0019773,GO:0006511
LH	chr2	95300001	95400000	g6359.t1	39S ribosomal protein L32, mitochondrial, partial	GO:0005762,GO:0003735,GO:0006412
LH	chr2	95800001	95900000	g6366.t1	biliverdin reductase A	GO:0005737,GO:0000166,GO:0004074, GO:0008270,GO:0042167
LH	chr2	99500001	99600000	g6469.t1	LOW QUALITY PROTEIN: tensin-3	
LH	chr2	99500001	99600000	g6470.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	
LH	chr2	99500001	99600000	g6471.t1	LOW QUALITY PROTEIN: tensin-3	GO:0016791,GO:0016311
LH	chr2	99500001	99600000	g6472.t1	hypothetical protein, partial	

LH	chr2	1.33E+08	1.34E+08	g7037.t1	protein MID1-COMPLEMENTING ACTIVITY 1-like	
LH	chr2	1.34E+08	1.34E+08	g7039.t1	hypothetical protein CAPTEDRAFT_223627	GO:0004222,GO:0008270,GO:0006508
LH	chr2	1.34E+08	1.34E+08	g7045.t1	probable G-protein coupled receptor 158 isoform X2	GO:0005886,GO:0016021,GO:0004930, GO:0007186,GO:0072659
LH	chr2	1.35E+08	1.36E+08	g7071.t1	PREDICTED: BMP and activin membrane-bound inhibitor homolog	GO:0008360,GO:0032092,GO:0008284, GO:0090263,GO:0005886,GO:0045668, GO:0030514,GO:0010718,GO:0005109, GO:0045893,GO:0005737,GO:0016021, GO:0016477,GO:0030512
LH	chr3	90400001	90500000	g8934.t1	protein FAM83B	GO:0005737,GO:0016020,GO:0005154, GO:0019901,GO:0036312,GO:0036313, GO:0007173,GO:0008283
LH	chr3	90400001	90500000	g8935.t1	hypothetical protein AAES_11399	GO:0005737,GO:0016020,GO:0005154, GO:0019901,GO:0036312,GO:0036313, GO:0007173,GO:0008283
LH	chr3	90500001	90600000	g8936.t1	PREDICTED: uncharacterized protein LOC106029472	GO:0006811
LH	chr3	90500001	90600000	g8937.t1	-	
LH	chr3	90500001	90600000	g8938.t1	hypothetical protein llap_1659	GO:0003964,GO:0006278
LH	chr3	90600001	90700000	g8940.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
LH	chr3	94100001	94200000	g8999.t1	hypothetical protein N308_05525, partial	
LH	chr3	94700001	94800000	g9005.t1	protein FAM110C, partial	
LH	chr3	94900001	95000000	g9007.t1	PREDICTED: low molecular weight phosphotyrosine protein phosphatase isoform X1	GO:0005829,GO:0009898,GO:0042383, GO:0003993,GO:0004726,GO:0035335
LH	chr3	94900001	95000000	g9008.t1	Protein FAM150B, partial	GO:0005576,GO:0030298,GO:0030971, GO:0061098
LH	chr3	95100001	95200000	g9011.t1	PREDICTED: transmembrane protein 18	GO:0016021,GO:0031965,GO:0003677, GO:0016477
LH	chr3	95300001	95400000	g9013.t1	rna-directed dna polymerase from mobile element jockey-like	

LH	chr3	95900001	96000000	g9019.t1	hypothetical protein	GO:0003964,GO:0006278
LH	chr3	95900001	96000000	g9020.t1	Protein tilB	
LH	chr3	95900001	96000000	g9021.t1	hypothetical protein N340_08264, partial	
LH	chr3	96100001	96200000	g9026.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
LH	chr3	96400001	96500000	g9033.t1	Protein tyrosine phosphatase-like protein PTPLAD2	GO:0005789,GO:0016021,GO:0003677, GO:0005524,GO:0102158,GO:0102343, GO:0102344,GO:0102345,GO:0006633, GO:0032508
LH	chr3	96400001	96500000	g9034.t1	hypothetical protein llap_14200	
LH	chr3	96800001	96900000	g9042.t1	hypothetical protein llap_8520	GO:0016021
LH	chr3	96800001	96900000	g9043.t1	trafficking protein particle complex subunit 12	GO:0000776,GO:0005654,GO:0005793, GO:0005794,GO:0030008,GO:0048471, GO:0006888,GO:0007030,GO:0051310, GO:0090234,GO:1905342
LH	chr3	96800001	96900000	g9044.t1	Ribonuclease H1	GO:0005634,GO:0005737,GO:0005886, GO:0000287,GO:0003676,GO:0004523, GO:0005506,GO:0010309,GO:0019509, GO:0090502
LH	chr3	98300001	98400000	g9070.t1	hypothetical protein llap_19307	GO:0003964,GO:0006278
LH	chr3	98300001	98400000	g9071.t1	ubiquitin carboxyl-terminal hydrolase 4	GO:0016787
LH	chr3	98300001	98400000	g9072.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0016021,GO:0003964,GO:0006278
LH	chr3	98900001	99000000	g9081.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003964,GO:0006278
LH	chr3	99200001	99300000	g9084.t1	Membrane-bound O-acyltransferase domain-containing protein 2, partial	GO:0016021,GO:0016746
LH	chr4	48400001	48500000	g12447.t1	palladin isoform X1	GO:0016477,GO:0030036
LH	chr4	48400001	48500000	g12448.t1	palladin isoform X1	GO:0016477,GO:0030036
LH	chr5	15600001	15700000	g13136.t1	golgin subfamily A member 5 isoform X1	GO:0000139,GO:0016021,GO:0007030
LH	chr5	15600001	15700000	g13137.t1	legumain	GO:0110165,GO:0004197,GO:0051603
LH	chr5	54100001	54200000	g13932.t1	hypothetical protein	

LH	chr5	54100001	54200000	g13933.t1	Coiled-coil domain-containing protein 34, partial	
LH	chr7	16300001	16400000	g14286.t1	PREDICTED: secretory phospholipase A2 receptor	GO:0016021,GO:0030246,GO:0006897
LH	chr6	9000001	9100000	g14788.t1	protein FAM35A isoform X1	GO:0005634,GO:0035861,GO:0045830, GO:2000042,GO:2001034
LH	chr6	9000001	9100000	g14789.t1	G protein-regulated inducer of neurite outgrowth 2	GO:0005886,GO:0031175
LH	chr6	9000001	9100000	g14790.t1	nuclear receptor coactivator 4 isoform X1	GO:0044754,GO:0003713,GO:0006622, GO:0006879,GO:0009725,GO:0045893
LH	chr6	9000001	9100000	g14791.t1	PREDICTED: mitochondrial import inner membrane translocase subunit Tim23	GO:0005744,GO:0005758,GO:0031305, GO:0008320,GO:0030150
LH	chr6	10800001	10900000	g14814.t1	S-adenosylmethionine synthase	GO:0004478,GO:0005524,GO:0046872, GO:0006556,GO:0006730
LH	chr6	10800001	10900000	g14815.t1	uncharacterized protein LOC101798065	
LH	chr6	10800001	10900000	g14816.t1	pulmonary surfactant-associated protein A-like isoform X1	
LH	chr6	10800001	10900000	g14817.t1	pulmonary surfactant-associated protein A	GO:0005581
LH	chr6	11400001	11500000	g14820.t1	nitroreductase	
LH	chr6	11500001	11600000	g14821.t1	ribosome biogenesis protein BMS1 homolog	GO:0005730,GO:0005525,GO:0042254
LH	chr6	11500001	11600000	g14822.t1	PREDICTED: uncharacterized protein LOC104015441	GO:0005634,GO:0003677
LH	chr6	11500001	11600000	g14823.t1	transmembrane protein 254	GO:0016021
LH	chr6	11500001	11600000	g14824.t1	hypothetical protein ASZ78_003290	
LH	chr6	15500001	15600000	g14869.t1	bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 2	GO:0000139,GO:0016021,GO:0015016, GO:0019213,GO:0002002,GO:0002448, GO:0015014,GO:0030210
LH	chr6	15500001	15600000	g14870.t1	zinc finger SWIM domain-containing protein 8 isoform X2	GO:0008270
LH	chr6	15700001	15800000	g14875.t1	synaptopodin 2-like protein	GO:0005829,GO:0015629,GO:0016607, GO:0030018,GO:0030054,GO:0003779, GO:0032233

LH	chr6	15700001	15800000	g14876.t1	myozenin-1 isoform X1	GO:0030018
LH	chr6	15800001	15900000	g14877.t1	inactive ubiquitin carboxyl-terminal hydrolase 54 isoform X3	GO:0004843,GO:0016579
LH	chr6	17200001	17300000	g14906.t1	acyl-CoA desaturase	GO:0005730,GO:0005789,GO:0016021, GO:0004768,GO:0005506,GO:0006636 GO:0005783,GO:0005829,GO:0005887, GO:0009986,GO:0015629,GO:0031410, GO:0034704,GO:0043235,GO:0060170, GO:0097730,GO:0005262,GO:0005272, GO:0005509,GO:0015269,GO:0033040, GO:0042802,GO:0051371,GO:0001581, GO:0007224,GO:0009415,GO:0035725, GO:0051289,GO:0070588,GO:0071468, GO:0071805
LH	chr6	17200001	17300000	g14907.t1	polycystic kidney disease 2-like 1 protein isoform X1	GO:0005886,GO:0016021,GO:0031966, GO:0005524,GO:0008559,GO:0042908, GO:0055085
LH	chr6	17200001	17300000	g14908.t1	ATP-binding cassette sub-family G member 2-like	GO:0000930,GO:0005739,GO:0005829, GO:0031083,GO:0099078,GO:1904115, GO:0043015,GO:0008625,GO:0016197, GO:0032418,GO:0048490,GO:0097345
LH	chr6	17200001	17300000	g14909.t1	hypothetical protein H355_001440	GO:0016021
LH	chr6	18900001	19000000	g14937.t1	transmembrane protein 72 isoform X2	GO:0016788,GO:0016042
LH	chr6	18900001	19000000	g14938.t1	lipase member M isoform X1	GO:0016788,GO:0016042
LH	chr6	18900001	19000000	g14939.t1	Lipase member M, partial	GO:0016021
LH	chr6	18900001	19000000	g14940.t1	Ankyrin repeat domain-containing protein 22, partial	GO:0016788,GO:0006629
LH	chr6	18900001	19000000	g14941.t1	lipase member M	GO:0005829,GO:0004435,GO:0005085, GO:0007264,GO:0016042,GO:0050790
LH	chr6	19800001	19900000	g14968.t1	hypothetical protein ASZ78_013389, partial	GO:0016021,GO:0008467
LH	chr6	21200001	21300000	g14991.t1	heparan sulfate glucosamine 3-O-sulfotransferase 1	GO:0004674,GO:0004712,GO:0005524, GO:0106310,GO:0006468
LH	chr6	24000001	24100000	g15089.t1	STE20-like serine/threonine-protein kinase, partial	

LH	chr6	24000001	24100000	g15090.t1	Collagen alpha-1(XVII) chain	GO:0016020
LH	chr8	15000001	15100000	g15561.t1	Epithelial chloride channel protein, partial	GO:0016021,GO:0005229,GO:0008237, GO:0006508,GO:0006821
LH	chr8	15000001	15100000	g15562.t1	PREDICTED: endophilin-B1 isoform X7	GO:0000139,GO:0005741,GO:0008289, GO:0006915,GO:0061024
LH	chr8	15000001	15100000	g15563.t1	hypothetical protein H355_004347	
LH	chr8	18000001	18100000	g15637.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
LH	chr8	18000001	18100000	g15638.t1	reverse hypothetical protein	GO:0016020
LH	chr8	18500001	18600000	g15653.t1	BMP/retinoic acid-inducible neural-specific protein 3 isoform X2	GO:0030425,GO:0071300,GO:0005739, GO:0045930,GO:0045666,GO:0005783, GO:0043025,GO:0005576
LH	chr8	18500001	18600000	g15654.t1	rna-directed dna polymerase from mobile element jockey-like	
LH	chr8	18500001	18600000	g15655.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003676,GO:0003964, GO:0004523,GO:0046872,GO:0006278, GO:0090502
LH	chr8	18500001	18600000	g15656.t1	rna-directed dna polymerase from mobile element jockey-like	
LH	chr9	25100001	25200000	g16416.t1	hypothetical protein Anapl_03019	
LH	chr9	25100001	25200000	g16417.t1	hypothetical protein Z169_11379, partial	
LH	chr9	25200001	25300000	g16418.t1	profilin 2, isoform CRA_c, partial	GO:0016887,GO:0032781,GO:1900028, GO:0003785,GO:0098685,GO:0005546, GO:0098793,GO:0005856,GO:0098794, GO:0030837,GO:0005737,GO:0051496, GO:0098885,GO:0030838,GO:0010633, GO:2000300,GO:0016021,GO:0098978, GO:0050821,GO:0033138
LH	chr9	25200001	25300000	g16419.t1	PREDICTED: E3 ubiquitin-protein ligase RNF13 isoform X2	GO:0005654,GO:0005765,GO:0005789, GO:0005829,GO:0016021,GO:0031902, GO:0008432,GO:0046872,GO:0061630, GO:0006511,GO:0051865,GO:0070304
LH	chr9	25200001	25300000	g16420.t1	PREDICTED: COMM domain-containing	GO:0016021

					protein 2 isoform X4	
LH	chr9	25200001	25300000	g16421.t1	PREDICTED: WW domain-containing transcription regulator protein 1	GO:0035264,GO:0016604,GO:0008284, GO:0045669,GO:0045944,GO:0006469, GO:0003015,GO:0000122,GO:0010718, GO:0005667,GO:0005829,GO:0060390, GO:0017145,GO:0072307,GO:1900182, GO:0042803,GO:0060993,GO:0060271, GO:0045599,GO:0090090,GO:0001894, GO:0032835,GO:0003713,GO:0003714, GO:0031146,GO:0035329,GO:0016567
LH	chr9	25300001	25400000	g16422.t1	PREDICTED: pre-mRNA-processing factor 17	GO:0016772,GO:0044237
LH	chr9	25300001	25400000	g16423.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016020,GO:0003964,GO:0006278
LH	chr1 4	3200001	3300000	g18137.t1	Hepatocyte growth factor, partial	
LH	chr1 6	4600001	4700000	g19405.t1	Platelet glycoprotein Ib beta chain, partial	GO:1990779,GO:0042802,GO:0007597, GO:0010572,GO:0051209
LH	chr1 6	4600001	4700000	g19406.t1	PREDICTED: septin-5 isoform X3	GO:0005856,GO:0008021,GO:0005525, GO:0017157
LH	chr2 5	7100001	7200000	g22288.t1	PREDICTED: uncharacterized protein C11ORF34	
LH	chr2 5	7100001	7200000	g22289.t1	PREDICTED: 6-pyruvoyl tetrahydrobiopterin synthase	GO:0003874,GO:0046872,GO:0006729
LH	chr2 5	7100001	7200000	g22290.t1	beta,beta-carotene 9',10'-oxygenase isoform X1	GO:0005739,GO:0003834,GO:0010436, GO:0046872,GO:0102076,GO:0016121, GO:0016124,GO:0042574,GO:0051881, GO:2000377
LH	chr2 5	7100001	7200000	g22291.t1	interleukin-18	GO:0042119,GO:0014068,GO:1901224, GO:0042531,GO:0045944,GO:0042632, GO:0051092,GO:0048661,GO:0000165, GO:0006954,GO:0032725,GO:0001525, GO:0032729,GO:0051142,GO:0032740, GO:2000556,GO:0120162,GO:0070328,

						GO:0032148,GO:0035655,GO:0045630, GO:0042267,GO:0005125,GO:0045515, GO:0042104,GO:0005615,GO:0005737, GO:0032819,GO:0045662,GO:0031663, GO:0010744,GO:0071407,GO:0051897
LH	chr2 5	7100001	7200000	g22292.t1	succinate dehydrogenase [ubiquinone] cytochrome b small subunit, mitochondrial	GO:0005743,GO:0016021,GO:0046872, GO:0006099
LH	chr2 5	7100001	7200000	g22293.t1	uncharacterized protein C11orf57 homolog isoform X1	
LH	chr2 5	7100001	7200000	g22294.t1	PIH1 domain-containing protein 2	
LH	chrZ	51800001	51900000	g10995.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
LH	chrZ	51800001	51900000	g10996.t1	PREDICTED: thyroid peroxidase-like, partial	
LH	chrZ	51800001	51900000	g10997.t1	PREDICTED: probable global transcription activator SNF2L2, partial	GO:0016514,GO:0005524,GO:0016887, GO:0042393,GO:0140658,GO:0006338, GO:0006355,GO:0051276
LH	chrZ	52500001	52600000	g11010.t1	hypothetical protein llap_8938	
LH	chrZ	53300001	53400000	g11033.t1	hypothetical protein H355_012376, partial	GO:0005905,GO:0072583
LH	chrZ	53400001	53500000	g11035.t1	-	
LH	chrZ	53400001	53500000	g11036.t1	FAD binding domain-domain-containing protein	
LH	chrZ	54300001	54400000	g11055.t1	PREDICTED: uncharacterized protein LOC104910185	
LH	chrZ	54300001	54400000	g11056.t1	PREDICTED: 3-hydroxy-3-methylglutaryl-Coenzyme A reductase isoform X2	GO:0005778,GO:0005789,GO:0016021, GO:0004420,GO:0050661,GO:0006695, GO:0008299,GO:0015936
LH	chrZ	54300001	54400000	g11057.t1	PREDICTED: mediator of RNA polymerase II transcription subunit 18	GO:0016592,GO:0070847,GO:0003712, GO:0006357,GO:0006369
LH	chrZ	54500001	54600000	g11062.t1	-	
LH	chrZ	54500001	54600000	g11063.t1	solute carrier family 22 member 7-hypothetical protein	
LH	chrZ	54500001	54600000	g11064.t1	hypothetical protein AS28_00488, partial	

LH	chrZ	67800001	67900000	g11417.t1	RNA-directed DNA polymerase from mobile element jockey, partial	GO:0003964,GO:0006278
LH	chrZ	67800001	67900000	g11418.t1	TPA_asm: integrase	
LH	chrZ	67800001	67900000	g11419.t1	pol- hypothetical protein	
MH	chr1	11200001	11300000	g278.t1	hypothetical protein llap_11679	
MH	chr1	11200001	11300000	g279.t1	hypothetical protein N301_07803, partial	
MH	chr1	11200001	11300000	g280.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr1	11200001	11300000	g281.t1	hypothetical protein N323_03669, partial	
MH	chr1	11300001	11400000	g283.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr1	11300001	11400000	g284.t1	hypothetical protein N325_02030, partial	
MH	chr1	11400001	11500000	g285.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC104257901, partial	GO:0003964,GO:0006278
MH	chr1	11400001	11500000	g286.t1	hypothetical protein N308_12456, partial	
MH	chr1	11400001	11500000	g287.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	
MH	chr1	11700001	11800000	g295.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003964,GO:0006278
MH	chr1	11800001	11900000	g296.t1	hypothetical protein AV530_012697	
MH	chr1	11800001	11900000	g297.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr1	11800001	11900000	g298.t1	PREDICTED: enteropeptidase	
MH	chr1	11800001	11900000	g299.t1	hypothetical protein llap_7334	
MH	chr1	11900001	12000000	g300.t1	hypothetical protein llap_8938	
MH	chr1	11900001	12000000	g301.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0003964,GO:0006278
MH	chr1	11900001	12000000	g302.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
MH	chr1	11900001	12000000	g303.t1	PREDICTED: DEP domain-containing protein 1B, partial	

MH	chr1	12400001	12500000	g314.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr1	12400001	12500000	g315.t1	hypothetical protein llap_10261	
MH	chr1	12400001	12500000	g316.t1	T9SS C-terminal target domain-containing protein	
MH	chr1	12400001	12500000	g317.t1	hypothetical protein N335_05178, partial	
MH	chr1	12400001	12500000	g318.t1	rho GTPase-activating protein 42-like protein	GO:0005737,GO:0005096,GO:0007165, GO:0050790
MH	chr1	13200001	13300000	g333.t1	baculoviral IAP repeat-containing protein 2	GO:0005634,GO:0061630,GO:0046872, GO:0005737,GO:0043154,GO:0031398, GO:0051726,GO:0043027,GO:0060546
MH	chr1	13200001	13300000	g334.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr1	13200001	13300000	g335.t1	PREDICTED: porimin	GO:0016021
MH	chr1	13200001	13300000	g336.t1	matrilysin	GO:0031012,GO:0004222,GO:0008270, GO:0006508
MH	chr1	13200001	13300000	g337.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr1	13500001	13600000	g344.t1	hypothetical protein	GO:0003964,GO:0006278
MH	chr1	13900001	14000000	g351.t1	-	
MH	chr1	14500001	14600000	g362.t1	PREDICTED: glutamate receptor 4-like, partial	GO:0016021
MH	chr1	27000001	27100000	g598.t1	LOW QUALITY PROTEIN: retinoblastoma-associated protein	GO:0035189,GO:0030308,GO:0031175, GO:0000977,GO:2000134,GO:0048667, GO:0006325,GO:0000785,GO:0006357
MH	chr1	66700001	66800000	g1711.t1	PREDICTED: RAS protein activator like-3 isoform X2	GO:0003964,GO:0006278
MH	chr1	66700001	66800000	g1712.t1	PREDICTED: macrophage mannose receptor 1-like isoform X2	GO:0016301,GO:0016310
MH	chr1	66700001	66800000	g1713.t1	PREDICTED: neuropathy target esterase isoform X4	GO:0110165
MH	chr1	66700001	66800000	g1714.t1	LOW QUALITY PROTEIN: protein Wiz	
MH	chr1	66700001	66800000	g1715.t1	A-kinase anchor protein 8-like	GO:0003964,GO:0006278

MH	chr1	66700001	66800000	g1716.t1	PREDICTED: interleukin enhancer-binding factor 3 isoform X2	GO:0009897,GO:0043235,GO:0004896, GO:0019955,GO:0019221
MH	chr1	66700001	66800000	g1717.t1	PREDICTED: beta-1,3-galactosyltransferase 2-like	GO:0005743,GO:0016021,GO:0005471, GO:0140021,GO:1901029,GO:1990544
MH	chr1	66800001	66900000	g1719.t1	PREDICTED: transcription activator BRG1-like, partial	GO:0016021,GO:0015057,GO:0045028, GO:0007596,GO:0035589,GO:0070493
MH	chr1	66800001	66900000	g1720.t1	PREDICTED: transcription activator BRG1 isoform X7	GO:0003964,GO:0006278
MH	chr1	66800001	66900000	g1721.t1	PREDICTED: transcription activator BRG1 isoform X6	
MH	chr1	66800001	66900000	g1722.t1	long-chain fatty acid transport protein 1	
MH	chr1	66800001	66900000	g1723.t1	nucleoredoxin-like protein 1, partial	
MH	chr1	66800001	66900000	g1724.t1	GEM-interacting protein isoform X1	GO:0003723,GO:0016301,GO:0051018, GO:0016310,GO:0043484
MH	chr1	67000001	67100000	g1726.t1	PREDICTED: pre-B-cell leukemia transcription factor 4 isoform X2	
MH	chr1	67900001	68000000	g1742.t1	choline transporter-like protein 2	GO:0003824
MH	chr1	67900001	68000000	g1743.t1	PREDICTED: bromodomain-containing protein 4 isoform X2	
MH	chr1	1.57E+08	1.57E+08	g3713.t1	hypothetical protein DUI87_01111	GO:0003824
MH	chr1	1.57E+08	1.57E+08	g3714.t1	-	
MH	chr1	1.57E+08	1.57E+08	g3715.t1	hypothetical protein Anapl_17728	GO:0003964,GO:0006278
MH	chr1	1.9E+08	1.9E+08	g4353.t1	NUDIX hydrolase	
MH	chr2	3900001	4000000	g4539.t1	trafficking protein particle complex subunit 9-like	
MH	chr2	14300001	14400000	g4692.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0005737,GO:0003723,GO:0003964, GO:0006278
MH	chr2	14300001	14400000	g4693.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	
MH	chr2	14300001	14400000	g4694.t1	hypothetical protein AAES_27573	GO:0030414
MH	chr2	14300001	14400000	g4695.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
MH	chr2	14300001	14400000	g4696.t1	hypothetical protein	GO:0003964,GO:0006278

MH	chr2	14300001	14400000	g4697.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr2	14400001	14500000	g4699.t1	PREDICTED: uncharacterized protein LOC109279670	GO:0003676,GO:0004523,GO:0090502
MH	chr2	14400001	14500000	g4700.t1	hypothetical protein AS27_05525, partial	
MH	chr2	14400001	14500000	g4701.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr2	14400001	14500000	g4702.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC104290418	GO:0003676,GO:0004523,GO:0015074, GO:0090502
MH	chr2	14400001	14500000	g4703.t1	endogenous retrovirus group k member 19 pol	GO:0004190,GO:0006508
MH	chr2	14400001	14500000	g4704.t1	PREDICTED: uncharacterized protein LOC107055427	GO:0016772,GO:0044237
MH	chr2	14400001	14500000	g4705.t1	PREDICTED: carbonic anhydrase-related protein	GO:0005615,GO:0003964,GO:0004867, GO:0006278,GO:0010951
MH	chr2	14400001	14500000	g4706.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr2	14400001	14500000	g4707.t1	hypothetical protein N308_04312, partial	
MH	chr2	14500001	14600000	g4708.t1	hypothetical protein, partial	GO:0003964,GO:0006278
MH	chr2	14500001	14600000	g4709.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr2	14500001	14600000	g4710.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr2	14500001	14600000	g4711.t1	hypothetical protein AV530_012697	
MH	chr2	14500001	14600000	g4712.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
MH	chr2	50100001	50200000	g5434.t1	kinetochore protein NDC80 homolog	GO:0000940,GO:0005654,GO:0005737, GO:0005813,GO:0031262,GO:0030332, GO:0042802,GO:0140483,GO:0000132, GO:0007052,GO:0007057,GO:0008315, GO:0031647,GO:0051298,GO:0051301, GO:0051315,GO:0051383,GO:0090267, GO:1905342

MH	chr2	50500001	50600000	g5439.t1	hypothetical protein llap_9744	
MH	chr2	50500001	50600000	g5440.t1	hypothetical protein C5E43_29795, partial	GO:0003964,GO:0006278
MH	chr2	50500001	50600000	g5441.t1	lipoprotein	
MH	chr2	53100001	53200000	g5495.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
MH	chr2	53100001	53200000	g5496.t1	Roundabout-like protein 2	GO:0008046,GO:0016199,GO:0035385
MH	chr2	56300001	56400000	g5585.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0005488,GO:0140640,GO:0044260, GO:0090304
MH	chr2	56300001	56400000	g5586.t1	hypothetical protein llap_7622	
MH	chr2	56300001	56400000	g5587.t1	hypothetical protein, partial	
MH	chr2	56300001	56400000	g5588.t1	hypothetical protein	GO:0005743,GO:0034654
MH	chr2	56300001	56400000	g5589.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0003964,GO:0006278
MH	chr2	56300001	56400000	g5590.t1	hypothetical protein llap_1028	GO:0003964,GO:0006278
MH	chr2	56300001	56400000	g5591.t1	Retron-type reverse transcriptase	GO:0003964,GO:0006278
MH	chr2	56300001	56400000	g5592.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
MH	chr2	56300001	56400000	g5593.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0016740
MH	chr2	56400001	56500000	g5594.t1	hypothetical protein C8258_32110, partial	GO:0003964,GO:0005085,GO:0019208, GO:0006278,GO:0043087
MH	chr2	56400001	56500000	g5595.t1	Uncharacterised protein	GO:0003964,GO:0006278
MH	chr2	56400001	56500000	g5596.t1	hypothetical protein DUI87_16456	GO:0008076,GO:0005249,GO:0034765, GO:0071805
MH	chr2	56400001	56500000	g5597.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr2	56400001	56500000	g5598.t1	hypothetical protein llap_9314	
MH	chr2	56400001	56500000	g5599.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr2	56400001	56500000	g5600.t1	reverse hypothetical protein	
MH	chr2	56400001	56500000	g5601.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003964,GO:0006278

MH	chr2	77800001	77900000	g5970.t1	transmembrane protein 245 isoform X1	
MH	chr2	77800001	77900000	g5971.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0016301,GO:0006278, GO:0016310
MH	chr2	77800001	77900000	g5972.t1	PREDICTED: protein lunapark-B-like	GO:0005737,GO:0003723,GO:0003964, GO:0006278
MH	chr2	79800001	79900000	g6011.t1	hypothetical protein, partial	GO:0003964,GO:0006278
MH	chr2	79800001	79900000	g6012.t1	hypothetical protein C5E45_35180, partial	GO:0005634,GO:0007049,GO:0010468
MH	chr2	79800001	79900000	g6013.t1	reverse transcriptase family protein, partial	
MH	chr2	1.04E+08	1.04E+08	g6545.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0005215,GO:0016787,GO:0097159, GO:1901363,GO:0006812,GO:0009987, GO:0065007
MH	chr2	1.06E+08	1.06E+08	g6601.t1	hypothetical protein N308_05490, partial	
MH	chr2	1.06E+08	1.06E+08	g6602.t1	leucine-rich repeat-containing protein 14B	
MH	chr2	1.06E+08	1.06E+08	g6603.t1	reverse transcriptase family protein, partial	GO:0003824
MH	chr2	1.06E+08	1.06E+08	g6604.t1	leucine-rich repeat-containing protein 14B	
MH	chr2	1.47E+08	1.47E+08	g7253.t1	hypothetical protein H355_007674	GO:0005737,GO:0004820,GO:0005524, GO:0016740,GO:0006426
MH	chr2	1.47E+08	1.47E+08	g7254.t1	LOW QUALITY PROTEIN: zinc finger protein 208	
MH	chr3	26900001	27000000	g7850.t1	RAC-gamma serine/threonine-protein kinase isoform X3	GO:1905653,GO:1905564,GO:0048854, GO:0090050,GO:0004674,GO:0045766, GO:0000002,GO:0005524,GO:0004712, GO:0045793,GO:0048873,GO:0018105, GO:0032008,GO:0106310,GO:2000773
MH	chr3	79800001	79900000	g8745.t1	inactive serine protease 35	GO:0004252,GO:0006508
MH	chr3	79800001	79900000	g8746.t1	alpha-tocopherol transfer protein-like	
MH	chr3	1.01E+08	1.01E+08	g9109.t1	transcription factor E2F6 isoform X1	
MH	chr3	1.01E+08	1.01E+08	g9112.t1	PREDICTED: histidine-rich glycoprotein-like	
MH	chr3	1.02E+08	1.02E+08	g9121.t1	hypothetical protein	
MH	chr3	1.02E+08	1.02E+08	g9122.t1	rna-directed dna polymerase from mobile	GO:0003964,GO:0006278

					element jockey- hypothetical protein	
MH	chr3	1.02E+08	1.02E+08	g9123.t1	hypothetical protein llap_22310	
MH	chr3	1.02E+08	1.02E+08	g9131.t1	hypothetical protein llap_47	
MH	chr3	1.02E+08	1.02E+08	g9132.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016021,GO:0003964,GO:0002729, GO:0006278
MH	chr3	1.02E+08	1.02E+08	g9133.t1	hypothetical protein N339_04974, partial	
MH	chr3	1.03E+08	1.03E+08	g9152.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr3	1.08E+08	1.08E+08	g9270.t1	GC-rich sequence DNA-binding factor, partial	GO:0005634,GO:0003677,GO:0000398
MH	chr3	1.08E+08	1.08E+08	g9271.t1	39S ribosomal protein L19, mitochondrial, partial	GO:0005840,GO:0003735,GO:0006412
MH	chr4	35200001	35300000	g12297.t1	reverse transcriptase family protein	
MH	chr4	61600001	61700000	g12655.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr5	19600001	19700000	g13216.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	GO:0003964,GO:0006278
MH	chr5	19600001	19700000	g13217.t1	PREDICTED: uncharacterized protein LOC106885922	GO:0004190,GO:0006508
MH	chr5	20500001	20600000	g13230.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003964,GO:0006278
MH	chr5	20600001	20700000	g13232.t1	hypothetical protein, partial	GO:0003964,GO:0006278
MH	chr5	20600001	20700000	g13233.t1	hypothetical protein	GO:0003964,GO:0006278
MH	chr5	47900001	48000000	g13836.t1	MULTISPECIES: NDMA-dependent alcohol dehydrogenase	
MH	chr5	48000001	48100000	g13837.t1	hypothetical protein llap_9744	
MH	chr5	48000001	48100000	g13838.t1	hypothetical protein	
MH	chr5	48800001	48900000	g13844.t1	hypothetical protein llap_16894	
MH	chr7	11200001	11300000	g14206.t1	-	
MH	chr7	11600001	11700000	g14209.t1	PREDICTED: enteropeptidase	
MH	chr7	11700001	11800000	g14210.t1	hypothetical protein N322_06263, partial	
MH	chr7	18400001	18500000	g14305.t1	hypothetical protein Y1Q_0011559	GO:0001518,GO:0005244,GO:0005248,

						GO:0034765,GO:0035725
MH	chr6	33000001	33100000	g15212.t1	testis-expressed protein 36 isoform X1	
MH	chr6	33000001	33100000	g15213.t1	erythroid differentiation-related factor 1, partial	
MH	chr6	33000001	33100000	g15214.t1	BRCA2 and CDKN1A-interacting protein	GO:0005654,GO:0005813,GO:0005814, GO:0005829,GO:0019908,GO:0097431, GO:0019207,GO:0000079,GO:0006974, GO:0034453,GO:0061101,GO:0090307
MH	chr8	1	100000	g15252.t1	hypothetical protein CIB84_004672	GO:0005968,GO:0004663,GO:0046872, GO:0018344
MH	chr9	7900001	8000000	g16025.t1	adiponectin	GO:0005576,GO:0005581,GO:0005179, GO:0007165
MH	chr9	7900001	8000000	g16026.t1	PREDICTED: otospiralin	GO:0007605
MH	chr9	7900001	8000000	g16027.t1	PREDICTED: actin-related protein 6-like, partial	GO:0003964,GO:0016301,GO:0006278, GO:0016310
MH	chr9	12000001	12100000	g16165.t1	PREDICTED: MAP6 domain-containing protein 1	GO:0005856,GO:0016021,GO:0004252, GO:0005515,GO:0006508
MH	chr9	12000001	12100000	g16166.t1	LOW QUALITY PROTEIN: dual specificity phosphatase 28	GO:0004725,GO:0017017,GO:0017018, GO:0035335
MH	chr9	12000001	12100000	g16167.t1	coiled-coil domain-containing protein 81-like isoform X2	
MH	chr10	7900001	8000000	g17881.t1	Uncharacterized protein CXorf45, partial	GO:0004577,GO:0004843,GO:0006488, GO:0006508
MH	chr10	7900001	8000000	g17882.t1	PREDICTED: putative bifunctional UDP-N-acetylglucosamine transferase and deubiquitinase ALG13	GO:0004577,GO:0006488
MH	chr10	8000001	8100000	g17884.t1	neuronal migration protein doublecortin isoform X5	GO:0005737,GO:0042995,GO:0035556 GO:0019934,GO:0042802,GO:0032098, GO:0007187,GO:0007268,GO:0045600, GO:0007631,GO:0043397,GO:0045202, GO:0001662,GO:0051378,GO:0005887, GO:0030425,GO:0001587,GO:0007626,
MH	chr10	10000001	10100000	g17950.t1	5-hydroxytryptamine receptor 2C isoform X1	

						GO:0007208,GO:0051482,GO:0070374, GO:0098666,GO:0030594,GO:0071886, GO:0010513,GO:0051209,GO:0031644 GO:0016021,GO:0062023,GO:0005509,
MH	chr10	11100001	11200000	g17966.t1	PREDICTED: heparan-sulfate 6-O-sulfotransferase 2	GO:0017095,GO:0001569,GO:0015015, GO:0048048,GO:0060173
MH	chr10	12900001	13000000	g17999.t1	hypothetical protein	
MH	chr10	17700001	17800000	g18065.t1	hypothetical protein	
MH	chr14	18600001	18700000	g18480.t1	serine/threonine-protein kinase 32A isoform X2	GO:0004674,GO:0005524,GO:0018105, GO:0035556
MH	chr14	18600001	18700000	g18481.t1	PREDICTED: dihydropyrimidinase-related protein 3	GO:0005737,GO:0030426,GO:0016810, GO:0031005,GO:0010975,GO:0030334, GO:0048678,GO:0051017,GO:0051491
MH	chr21	9900001	10000000	g19197.t1	spermatogenesis-associated protein 2 isoform X1	
MH	chr21	9900001	10000000	g19198.t1	PREDICTED: E3 ubiquitin-protein ligase RNF114 isoform X1	GO:0005737,GO:0046872,GO:0061630, GO:0000209,GO:0006511,GO:0007283, GO:0030154
MH	chr21	9900001	10000000	g19199.t1	zinc finger protein SNAI1	GO:0000978,GO:0000981,GO:0006357
MH	chr21	9900001	10000000	g19200.t1	ubiquitin-conjugating enzyme E2 variant 1	GO:0005515,GO:0035370,GO:0000151, GO:0045893,GO:0006282,GO:0043123, GO:1902523,GO:0030154,GO:0005829, GO:0051092,GO:0006301,GO:0005654
MH	chr21	10000001	10100000	g19201.t1	PREDICTED: serine/arginine-rich splicing factor 6	GO:0000381,GO:0016607,GO:0006376, GO:0003729
MH	chr21	10000001	10100000	g19202.t1	-	
MH	chr21	10000001	10100000	g19203.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0003964,GO:0006278
MH	chr21	10000001	10100000	g19204.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile	GO:0003964,GO:0006278

					element jockey-like	
MH	chr2 1	10000001	10100000	g19205.t1	Lethal(3)malignant brain tumor-like protein, partial	GO:0005634,GO:0008270,GO:0006325, GO:0006355
MH	chr2 1	10000001	10100000	g19206.t1	PREDICTED: microtubule-associated proteins 1A/1B light chain 3A	GO:0000421,GO:0005829,GO:0044754, GO:0005543,GO:0008017,GO:0031625, GO:0000045,GO:0000422,GO:0006995, GO:0097352
MH	chr2 1	10000001	10100000	g19207.t1	PREDICTED: dynein light chain roadblock-type 1	GO:0005737,GO:0005868,GO:0005874, GO:0045505,GO:0007018
MH	chr2 9	4900001	5000000	g20980.t1	proline-rich protein 22	
MH	chr2 9	4900001	5000000	g20981.t1	tRNA-dihydrouridine synthase 3-like, partial	GO:0017150,GO:0046872,GO:0050660, GO:0002943
MH	chr2 9	4900001	5000000	g20982.t1	hypothetical protein Anapl_07859	
MH	chr2 9	4900001	5000000	g20983.t1	neurturin isoform X2	GO:0005576,GO:0016021,GO:0008083, GO:0030116,GO:0030971,GO:0007165, GO:0048731
MH	chr2 7	1	100000	g20989.t1	PREDICTED: FYVE, RhoGEF and PH domain-containing protein 2 isoform X9	
MH	chr2 7	1	100000	g20990.t1	PREDICTED: mitochondrial carrier homolog 1	GO:0016021
MH	chr2 7	1	100000	g20991.t1	-	
MH	chr2 7	100001	200000	g20993.t1	Putative protein C6orf89 like protein	GO:0006406,GO:0003723,GO:0043086, GO:1990830,GO:0005737,GO:0016607, GO:0048024,GO:0043274,GO:0051726, GO:0004860
MH	chr2 7	100001	200000	g20994.t1	PREDICTED: serine/threonine-protein kinase 38 isoform X1	GO:0004674,GO:0005524,GO:0006468
MH	chr2 7	100001	200000	g20995.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chr2 7	100001	200000	g20996.t1	BTB/POZ domain-containing protein KCTD20 isoform X1	GO:0005737

MH	chr2 7	100001	200000	g20997.t1	hypothetical protein llap_9902	GO:0003964,GO:0006278
MH	chr2 7	200001	300000	g20999.t1	Patatin-like phospholipase domain-containing protein 1, partial	GO:0016787,GO:0016042
MH	chr2 7	200001	300000	g21000.t1	hypothetical protein H355_010928	GO:0008076,GO:0005249,GO:0034765, GO:0051260,GO:0071805
MH	chr2 7	300001	400000	g21001.t1	potassium voltage-gated channel subfamily A member 2	GO:0008076,GO:0030425,GO:0032809, GO:0043194,GO:0043204,GO:0043679, GO:0044224,GO:0005251,GO:0014059, GO:0021633,GO:0034765,GO:0045188, GO:0051260,GO:0071805
MH	chr2 7	300001	400000	g21002.t1	potassium voltage-gated channel subfamily A member 10	GO:0008076,GO:0005249,GO:0034765, GO:0051260,GO:0071805
MH	chr2 7	300001	400000	g21003.t1	Embryonic pepsinogen, partial	GO:0004190,GO:0006508,GO:0007586
MH	chr2 7	300001	400000	g21004.t1	prokineticin-1	GO:0005576
MH	chr2 7	300001	400000	g21005.t1	LOW QUALITY PROTEIN: monocarboxylate transporter 5	GO:0005887,GO:0008028,GO:0015718, GO:0055085
MH	chr2 7	300001	400000	g21006.t1	putative RNA-binding protein 15	GO:0016607,GO:0036396,GO:0003729, GO:0000122,GO:0000381,GO:0001510, GO:0001569,GO:0007221,GO:0038163, GO:0045638,GO:0045652,GO:0048536, GO:0060412
MH	chr2 2	3900001	4000000	g22003.t1	PREDICTED: calsyntenin-1 isoform X2	GO:0000139,GO:0005783,GO:0005886, GO:0016021,GO:0042995,GO:0045202, GO:0001540,GO:0005509,GO:0019894, GO:0042988,GO:0007156
MH	chr2 2	3900001	4000000	g22004.t1	LOW QUALITY PROTEIN: phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform	GO:0014065,GO:1905278,GO:0038089, GO:0033031,GO:0036092,GO:0046934, GO:0005886,GO:0005943,GO:0010628, GO:0005829,GO:0001938,GO:0016310, GO:0052812,GO:0010595,GO:0035005, GO:0051897,GO:0016303

MH	chr2 4	700001	800000	g22357.t1	hypothetical protein H355_000517	GO:0005764,GO:0004419,GO:0004560, GO:0006004,GO:0006629,GO:0016139
MH	chr2 4	700001	800000	g22358.t1	cannabinoid receptor 2	GO:0005886,GO:0016021,GO:0003676, GO:0004949,GO:0038171
MH	chr2 4	700001	800000	g22359.t1	basic fatty acid binding protein	GO:0005634,GO:0005829,GO:0005504, GO:0015908
MH	chr2 4	700001	800000	g22360.t1	PREDICTED: LOW QUALITY PROTEIN: myomesin-3	
MH	chr2 4	700001	800000	g22361.t1	myomesin-3	
MH	chr2 4	700001	800000	g22362.t1	interleukin-22 receptor subunit alpha-1	GO:0016021
MH	chrZ	25000001	25100000	g10113.t1	hypothetical protein	GO:0016020,GO:0004888,GO:0007165
MH	chrZ	25000001	25100000	g10114.t1	adaptin ear-binding coat-associated protein 1	GO:0003964,GO:0006278
MH	chrZ	25000001	25100000	g10115.t1	hypothetical protein	
MH	chrZ	25000001	25100000	g10116.t1	hypothetical protein AS27_13285, partial	GO:0003964,GO:0006278
MH	chrZ	25000001	25100000	g10117.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0003964,GO:0006278
MH	chrZ	26200001	26300000	g10162.t1	hypothetical protein llap_15528	
MH	chrZ	57200001	57300000	g11124.t1	hypothetical protein C5F51_36540, partial	GO:0003964,GO:0006278
MH	chrZ	57200001	57300000	g11125.t1	hypothetical protein	
MH	chrZ	58200001	58300000	g11154.t1	rna-directed dna polymerase from mobile element jockey-like	
MH	chrZ	58200001	58300000	g11155.t1	hypothetical protein AV530_012697	
MH	chrZ	58300001	58400000	g11157.t1	LOW QUALITY PROTEIN: E3 ubiquitin-protein ligase RNF180	
MH	chrZ	58300001	58400000	g11158.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein	
MH	chrZ	58300001	58400000	g11159.t1	PREDICTED: 5-hydroxytryptamine receptor 1A	
MH	chrZ	58700001	58800000	g11164.t1	rna-directed dna polymerase from mobile element jockey-like	

MH	chrZ	58700001	58800000	g11165.t1	hypothetical protein N340_05374, partial
MH	chrZ	58700001	58800000	g11166.t1	glycosyltransferase family 1 protein
MH	chrZ	58700001	58800000	g11167.t1	rna-directed dna polymerase from mobile element jockey-like
MH	chrZ	65300001	65400000	g11352.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
MH	chrZ	65300001	65400000	g11353.t1	PI-PLC X domain-containing protein 3
MH	chrZ	65300001	65400000	g11354.t1	hypothetical protein llap_2283
MH	chrZ	65300001	65400000	g11355.t1	PREDICTED: PI-PLC X domain-containing protein 3
PH	chr1	12100001	12200000	g306.t1	reverse hypothetical protein
PH	chr1	12100001	12200000	g307.t1	PREDICTED: inositol 1,4,5-trisphosphate receptor-interacting protein-like 1
PH	chr1	12100001	12200000	g308.t1	PREDICTED: thyroid peroxidase-like, partial
PH	chr1	12100001	12200000	g309.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	32700001	32800000	g733.t1	dtw domain-containing protein 2
PH	chr1	32700001	32800000	g734.t1	ryanodine receptor hypothetical protein
PH	chr1	32700001	32800000	g735.t1	hypothetical protein llap_2032
PH	chr1	32700001	32800000	g736.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr1	32700001	32800000	g737.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	33400001	33500000	g746.t1	hypothetical protein
PH	chr1	33400001	33500000	g747.t1	methylase
PH	chr1	33400001	33500000	g748.t1	reverse hypothetical protein
PH	chr1	33800001	33900000	g755.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	33800001	33900000	g756.t1	protocadherin-20
PH	chr1	33800001	33900000	g757.t1	rna-directed dna polymerase from mobile

					element jockey-like
PH	chr1	33800001	33900000	g758.t1	hypothetical protein HGM15179_020528
PH	chr1	33800001	33900000	g759.t1	rna-directed dna polymerase from mobile element hypothetical protein
PH	chr1	33900001	34000000	g760.t1	hypothetical protein llap_2565
PH	chr1	33900001	34000000	g761.t1	hypothetical protein
PH	chr1	34000001	34100000	g762.t1	hypothetical protein llap_9744
PH	chr1	34000001	34100000	g763.t1	hypothetical protein C8258_32110, partial
PH	chr1	34000001	34100000	g764.t1	PREDICTED: uncharacterized protein LOC106047159
PH	chr1	34000001	34100000	g765.t1	hypothetical protein DUI87_13438
PH	chr1	34000001	34100000	g766.t1	uncharacterized protein LOC112531750
PH	chr1	34000001	34100000	g767.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	34100001	34200000	g768.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	34100001	34200000	g769.t1	hypothetical protein DUI87_25616
PH	chr1	34100001	34200000	g770.t1	hypothetical protein llap_3836
PH	chr1	35900001	36000000	g859.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	35900001	36000000	g860.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	35900001	36000000	g861.t1	hypothetical protein DUI87_24024
PH	chr1	37400001	37500000	g909.t1	hypothetical protein
PH	chr1	37400001	37500000	g910.t1	vps10 domain-containing receptor 2
PH	chr1	37500001	37600000	g911.t1	hypothetical protein DUI87_13490
PH	chr1	37500001	37600000	g912.t1	-
PH	chr1	37500001	37600000	g913.t1	-
PH	chr1	37500001	37600000	g914.t1	pol- hypothetical protein
PH	chr1	37500001	37600000	g915.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein

PH	chr1	37500001	37600000	g916.t1	PREDICTED: VPS10 domain-containing receptor SorCS2
PH	chr1	37500001	37600000	g917.t1	coiled-coil domain-containing protein 70-like
PH	chr1	37800001	37900000	g924.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr1	37800001	37900000	g925.t1	reverse transcriptase family protein, partial
PH	chr1	37800001	37900000	g926.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr1	39000001	39100000	g973.t1	protein eva-1 homolog A isoform X1
PH	chr1	39000001	39100000	g974.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein
PH	chr1	79600001	79700000	g2006.t1	transmembrane protein 201, partial
PH	chr1	79600001	79700000	g2007.t1	PREDICTED: solute carrier family 25 member 33 isoform X1
PH	chr1	85000001	85100000	g2139.t1	Signal recognition particle receptor subunit alpha, partial
PH	chr1	85000001	85100000	g2140.t1	PREDICTED: protein FAM118B isoform X2
PH	chr1	85000001	85100000	g2141.t1	Toll/interleukin-1 receptor domain-containing adapter protein, partial
PH	chr1	1.19E+08	1.19E+08	g2771.t1	solute carrier family 2, facilitated glucose transporter member 3 isoform X1
PH	chr1	1.19E+08	1.19E+08	g2772.t1	PREDICTED: uncharacterized protein LOC108997871
PH	chr1	1.19E+08	1.19E+08	g2773.t1	homeobox protein NANOG
PH	chr1	1.19E+08	1.19E+08	g2774.t1	homeobox protein NANOG
PH	chr1	1.7E+08	1.7E+08	g3922.t1	hypothetical protein HGM15179_019125
PH	chr1	1.7E+08	1.7E+08	g3923.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial
PH	chr1	1.82E+08	1.82E+08	g4193.t1	coiled-coil domain-containing protein 71L
PH	chr2	400001	500000	g4475.t1	reverse transcriptase family protein, partial

PH	chr2	400001	500000	g4476.t1	protein FAM83H
PH	chr2	400001	500000	g4477.t1	PREDICTED: mitogen-activated protein kinase 15 isoform X1
PH	chr2	400001	500000	g4478.t1	coiled-coil domain-containing protein 166
PH	chr2	64300001	64400000	g5760.t1	reverse hypothetical protein
PH	chr2	64700001	64800000	g5762.t1	zinc finger protein, partial
PH	chr2	64700001	64800000	g5763.t1	hypothetical protein N322_06263, partial
PH	chr2	78900001	79000000	g5988.t1	hypothetical protein DUI87_30491
PH	chr2	78900001	79000000	g5989.t1	reverse hypothetical protein
PH	chr2	78900001	79000000	g5990.t1	reverse transcriptase, partial
PH	chr2	78900001	79000000	g5991.t1	hypothetical protein DUI87_19212
PH	chr2	79300001	79400000	g5999.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr2	79300001	79400000	g6000.t1	rna-directed dna polymerase from mobile element jockey- hypothetical protein
PH	chr2	79300001	79400000	g6001.t1	hypothetical protein
PH	chr2	79300001	79400000	g6002.t1	hypothetical protein llap_3944
PH	chr2	79600001	79700000	g6005.t1	cadherin-10-like
PH	chr2	79600001	79700000	g6006.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr2	79600001	79700000	g6007.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr2	79600001	79700000	g6008.t1	hypothetical protein llap_6671
PH	chr2	84200001	84300000	g6085.t1	hypothetical protein AV530_016553
PH	chr2	84200001	84300000	g6086.t1	pol- hypothetical protein
PH	chr2	1.02E+08	1.02E+08	g6517.t1	E3 ubiquitin-protein ligase RNF144B, partial
PH	chr2	1.12E+08	1.12E+08	g6700.t1	hypothetical protein llap_2921
PH	chr2	1.12E+08	1.12E+08	g6701.t1	3-oxoacyl-[acyl-carrier-protein] synthase, mitochondrial

PH	chr2	1.33E+08	1.33E+08	g7026.t1	dnaJ homolog subfamily C member 1
PH	chr2	1.34E+08	1.34E+08	g7039.t1	hypothetical protein CAPTEDRAFT_223627
PH	chr2	1.34E+08	1.34E+08	g7046.t1	Putative G-protein coupled receptor 158, partial
PH	chr2	1.34E+08	1.34E+08	g7047.t1	hypothetical protein
PH	chr2	1.35E+08	1.35E+08	g7057.t1	acyl-CoA-binding domain-containing protein 5 isoform X1
PH	chr2	1.35E+08	1.35E+08	g7058.t1	LOW QUALITY PROTEIN: serine/threonine-protein kinase greatwall
PH	chr2	1.35E+08	1.35E+08	g7059.t1	hypothetical protein ASZ78_000070
PH	chr2	1.35E+08	1.35E+08	g7060.t1	LOW QUALITY PROTEIN: patched domain-containing protein 3
PH	chr2	1.36E+08	1.36E+08	g7072.t1	hypothetical protein
PH	chr2	1.36E+08	1.36E+08	g7073.t1	-
PH	chr2	1.47E+08	1.47E+08	g7251.t1	villin-like protein isoform X1
PH	chr2	1.47E+08	1.47E+08	g7252.t1	-
PH	chr3	45300001	45400000	g8123.t1	parkin coregulated gene protein isoform X2
PH	chr3	57500001	57600000	g8322.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr3	57500001	57600000	g8323.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC104460923, partial
PH	chr3	57500001	57600000	g8324.t1	dtw domain-containing protein 2
PH	chr3	90300001	90400000	g8931.t1	orexin receptor type 2 isoform X1
PH	chr3	90300001	90400000	g8932.t1	RNA-directed DNA polymerase from mobile element jockey, partial
PH	chr3	90300001	90400000	g8933.t1	rna-directed dna polymerase from mobile element hypothetical protein
PH	chr3	91300001	91400000	g8956.t1	PREDICTED: angiopoietin-2 isoform X1
PH	chr3	91500001	91600000	g8958.t1	hypothetical protein, partial

PH	chr3	92400001	92500000	g8971.t1	PREDICTED: enteropeptidase
PH	chr3	92400001	92500000	g8972.t1	hypothetical protein N334_14356, partial
PH	chr3	92400001	92500000	g8973.t1	hypothetical protein
PH	chr3	92400001	92500000	g8974.t1	hypothetical protein AS27_13285, partial
PH	chr3	92500001	92600000	g8975.t1	hypothetical protein
PH	chr3	95000001	95100000	g9009.t1	-
PH	chr3	95000001	95100000	g9010.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	95200001	95300000	g9012.t1	hypothetical protein llap_2362
PH	chr3	95300001	95400000	g9013.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	96300001	96400000	g9029.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	96300001	96400000	g9030.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial
PH	chr3	96300001	96400000	g9031.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	96700001	96800000	g9039.t1	hypothetical protein AS28_13382, partial
PH	chr3	96700001	96800000	g9040.t1	PREDICTED: protein TSSC1-like, partial
PH	chr3	97100001	97200000	g9048.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	97200001	97300000	g9049.t1	hypothetical protein llap_17364
PH	chr3	97200001	97300000	g9050.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr3	97200001	97300000	g9051.t1	hypothetical protein
PH	chr3	99200001	99300000	g9084.t1	Membrane-bound O-acyltransferase domain-containing protein 2, partial
PH	chr3	1.12E+08	1.13E+08	g9354.t1	hypothetical protein UY3_10755
PH	chr4	8600001	8700000	g11855.t1	nucleolar protein 14 isoform A
PH	chr4	30900001	31000000	g12246.t1	inactive ubiquitin carboxyl-terminal hydrolase 53
PH	chr4	31100001	31200000	g12248.t1	60S ribosomal protein L7-like 1

PH	chr4	31100001	31200000	g12249.t1	Mitotic spindle assembly checkpoint protein MAD2A
PH	chr4	36500001	36600000	g12309.t1	putative RNA-directed DNA polymerase from transposon X-element
PH	chr4	36600001	36700000	g12311.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr4	36600001	36700000	g12312.t1	rna-directed dna polymerase from mobile element jockey-like
PH	chr4	48300001	48400000	g12444.t1	annexin A10
PH	chr4	48300001	48400000	g12445.t1	serine-rich adhesin for platelets-like isoform X1
PH	chr4	48600001	48700000	g12451.t1	carbonyl reductase family member 4 isoform X1
PH	chr4	48600001	48700000	g12452.t1	SH3R1 ligase
PH	chr5	15600001	15700000	g13136.t1	golgin subfamily A member 5 isoform X1
PH	chr5	15600001	15700000	g13137.t1	legumain
PH	chr5	19800001	19900000	g13220.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr5	19900001	20000000	g13221.t1	hypothetical protein llap_9744
PH	chr5	35100001	35200000	g13637.t1	PREDICTED: coiled-coil domain-containing protein 177
PH	chr5	35100001	35200000	g13638.t1	pleckstrin homology domain-containing family D member 1
PH	chr7	12400001	12500000	g14212.t1	RNA-binding protein 4F, partial
PH	chr7	16200001	16300000	g14284.t1	unconventional myosin-X-like, partial
PH	chr6	8600001	8700000	g14782.t1	PREDICTED: melanopsin isoform X1
PH	chr6	11500001	11600000	g14821.t1	ribosome biogenesis protein BMS1 homolog
PH	chr6	11500001	11600000	g14822.t1	PREDICTED: uncharacterized protein LOC104015441
PH	chr6	11500001	11600000	g14823.t1	transmembrane protein 254

PH	chr6	11500001	11600000	g14824.t1	hypothetical protein ASZ78_003290
PH	chr9	17800001	17900000	g16294.t1	coiled-coil domain-containing protein 39
PH	chr9	17800001	17900000	g16295.t1	PREDICTED: tetratricopeptide repeat protein 14
PH	chr9	20400001	20500000	g16322.t1	PREDICTED: phospholipase D1 isoform X3
PH	chr1 1	9700001	9800000	g17032.t1	zinc finger protein basonuclin-1 isoform X1
PH	chr1 1	16200001	16300000	g17186.t1	probable threonine--tRNA ligase 2, cytoplasmic
PH	chr1 1	16200001	16300000	g17187.t1	hypothetical protein H355_004157
PH	chr1 1	16200001	16300000	g17188.t1	PREDICTED: adenosine deaminase-like protein isoform X2
PH	chr1 1	16200001	16300000	g17189.t1	la-related protein 6 isoform X2
PH	chr1 1	16200001	16300000	g17190.t1	leucine-rich repeat-containing protein 49
PH	chr1 4	3000001	3100000	g18136.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like
PH	chr1 5	13500001	13600000	g18863.t1	hypothetical protein
PH	chr1 5	13500001	13600000	g18864.t1	HAMP domain-containing protein
PH	chr1 5	16200001	16300000	g18919.t1	cytosolic Fe-S cluster assembly factor NUBP1
PH	chr1 5	16200001	16300000	g18920.t1	Golgi apparatus membrane protein TVP23 homolog A, partial
PH	chr1 5	16200001	16300000	g18921.t1	hypothetical protein, partial
PH	chr1 5	16200001	16300000	g18922.t1	hypothetical protein
PH	chr1	16200001	16300000	g18923.t1	LOW QUALITY PROTEIN: MHC class II

	5				transactivator
PH	chr1 5	16200001	16300000	g18924.t1	PREDICTED: dexamethasone-induced protein
PH	chr1 8	5400001	5500000	g20256.t1	Ectonucleoside triphosphate diphosphohydrolase 2, partial
PH	chr1 8	5400001	5500000	g20257.t1	hypothetical protein
PH	chr1 8	5400001	5500000	g20258.t1	PREDICTED: protein PAXX isoform X3
PH	chr1 8	5400001	5500000	g20259.t1	ATP-binding cassette sub-family A member 2
PH	chr2 5	6600001	6700000	g22281.t1	T-cell surface glycoprotein CD3 delta chain
PH	chr2 5	6600001	6700000	g22282.t1	PREDICTED: centromere/kinetochore protein zw10 homolog
PH	chr2 5	6600001	6700000	g22283.t1	transmembrane protease serine 5 isoform X7
PH	chr2 5	6600001	6700000	g22284.t1	D(2) dopamine receptor isoform X2
PH	chr2 5	6700001	6800000	g22286.t1	neural cell adhesion molecule 1 isoform X2
PH	chrZ	52500001	52600000	g11010.t1	hypothetical protein llap_8938
PH	chrZ	53400001	53500000	g11035.t1	-
PH	chrZ	53400001	53500000	g11036.t1	FAD binding domain-domain-containing protein
PH	chrZ	53500001	53600000	g11037.t1	Transcription factor BTF3, partial
PH	chrZ	53500001	53600000	g11038.t1	Ankyrin repeat family A protein 2
PH	chrZ	54000001	54100000	g11047.t1	hypothetical protein CIB84_001443, partial
PH	chrZ	54000001	54100000	g11048.t1	beta-hexosaminidase subunit beta
PH	chrZ	67500001	67600000	g11410.t1	LOW QUALITY PROTEIN: NAD kinase 2, mitochondrial
PH	chrZ	67500001	67600000	g11411.t1	LOW QUALITY PROTEIN: S-phase kinase-associated protein 2

PH	chrZ	67500001	67600000	g11412.t1	LMBR1 domain-containing protein 2 isoform X6
PH	chrZ	67500001	67600000	g11413.t1	hypothetical protein llap_5660
PH	chrZ	67600001	67700000	g11414.t1	calcyphosin-like protein
PH	chrZ	67600001	67700000	g11415.t1	interleukin-7 receptor subunit alpha isoform X2

Table S1.7: GO Term enrichment analyses for 167 Autosome genes and 16 Z-linked genes under LH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be FDR- adjusted p values.

Categories:	DNA Information Processing	Cellular structure, metabolism and function	Behaviour	Individual's well-being and health (e.g.sleep,appetite, mood, energy)	Neurotransmission	Fertilization & embryonic development	Immune response

Windows Type	Chromosome Type	GO ID	Ontology	GO Name	Number of Genes	p-value	Adjusted p-value	Significance	Genes
LH	Autosomes	GO:0003964	mf	RNA-directed DNA polymerase activity	27	2.07E-02	0.0986	No	g674.t1,g748.t1,g790.t1,g791.t1,g795.t1,g798.t1,g806.t1,g807.t1,g818.t1,g819.t1,g820.t1,g823.t1,g826.t1,g828.t1,g831.t1,g925.t1,g3531.t1,g5304.t1,g6351.t1,g8938.t1,g9019.t1,g9026.t1,g9070.t1,g9072.t1,g9081.t1,g15655.t1,g16423.t1
LH	Autosomes	GO:0006278	bp	RNA-templated DNA biosynthetic process	27	2.08E-02	0.0986	No	g674.t1,g748.t1,g790.t1,g791.t1,g795.t1,g798.t1,g806.t1,g807.t1,g818.t1,g819.t1,g820.t1,g823.t1,g826.t1,g828.t1,g831.t1,g925.t1,g3531.t1,g5304.t1,g6351.t1,g8938.t1,g9019.t1,g9026.t1,g9070.t1,g9072.t1,g9081.t1,g15655.t1,g16423.t1

LH	Autosomes	GO:0016772	mf	transferase activity, transferring phosphorus-containing groups	3	0.03337	0.1159	No	g676.t1,g804.t1,g16422.t1
LH	Autosomes	GO:0044237	bp	cellular metabolic process	3	0.03687	0.1212	No	g676.t1,g804.t1,g16422.t1
LH	Autosomes	GO:0000276	cc	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	1	0.1075	0.2099	No	g746.t1
LH	Autosomes	GO:0015078	mf	proton transmembrane transporter activity	1	0.11383	0.2159	No	g746.t1
LH	Autosomes	GO:0006811	bp	monoatomic ion transport	2	0.03112	0.1159	No	g746.t1,g8936.t1
LH	Autosomes	GO:0015986	bp	proton motive force-driven ATP synthesis	1	0.12634	0.2267	No	g746.t1
LH	Autosomes	GO:0016301	mf	kinase activity	1	1	1	No	g791.t1
LH	Autosomes	GO:0016310	bp	phosphorylation	1	1	1	No	g791.t1
LH	Autosomes	GO:0003824	mf	catalytic activity	7	0.09288	0.199	No	g799.t1,g821.t1,g924.t1,g959.t1,g2141.t1,g8940.t1,g15637.t1
LH	Autosomes	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	26	1	1	No	g820.t1,g831.t1,g3031.t1,g7045.t1,g7071.t1,g9011.t1,g9033.t1,g9042.t1,g9072.t1,g9081.t1,g9084.t1,g13136.t1,g14286.t1,g14823.t1,g14869.t1,g14906.t1,g14908.t1,g14937.t1,g14940.t1,g14991.t1,g15561.t1,g15655.t1,g16418.t1,g16419.t1,g16420.t1,g22292.t1
LH	Autosomes	GO:0030234	mf	enzyme regulator activity	2	0.00369	0.0712	No	g821.t1,g926.t1

LH	Autosomes	GO:0016020	cc	membrane	6	0.33148	0.4249	No	g823.t1,g8934.t1,g8935.t1,g15090.t1,g15638.t1,g16423.t1
LH	Autosomes	GO:0007165	bp	signal transduction	1	0.5276	0.6082	No	g823.t1
LH	Autosomes	GO:0004867	mf	serine-type endopeptidase inhibitor activity	2	0.17809	0.2675	No	g959.t1,g6351.t1
LH	Autosomes	GO:0010951	bp	negative regulation of endopeptidase activity	3	0.05817	0.1533	No	g959.t1,g5456.t1,g6351.t1
LH	Autosomes	GO:0004190	mf	aspartic-type endopeptidase activity	1	0.23135	0.3214	No	g960.t1
LH	Autosomes	GO:0006508	bp	proteolysis	3	0.74835	0.8067	No	g960.t1,g7039.t1,g15561.t1
LH	Autosomes	GO:0005654	cc	nucleoplasm	3	0.63729	0.7076	No	g3033.t1,g9043.t1,g16419.t1
LH	Autosomes	GO:0005737	cc	cytoplasm	10	1	1	No	g3033.t1,g3792.t1,g5303.t1,g6366.t1,g7071.t1,g8934.t1,g8935.t1,g9044.t1,g16418.t1,g22291.t1
LH	Autosomes	GO:0008432	mf	JUN kinase binding	2	0.00015	0.0165	**Yes**	g3033.t1,g16419.t1
LH	Autosomes	GO:0017017	mf	MAP kinase tyrosine/serine/threonine phosphatase activity	1	0.08826	0.1954	No	g3033.t1
LH	Autosomes	GO:0017018	mf	myosin phosphatase activity	1	0.25823	0.3443	No	g3033.t1
LH	Autosomes	GO:0048273	mf	mitogen-activated protein kinase p38 binding	1	0.01411	0.0919	No	g3033.t1
LH	Autosomes	GO:0006470	bp	protein dephosphorylation	1	0.34747	0.442	No	g3033.t1
LH	Autosomes	GO:0043508	bp	negative regulation of JUN kinase activity	1	0.01411	0.0919	No	g3033.t1
LH	Autosomes	GO:0045204	bp	MAPK export from nucleus	1	0.00708	0.0712	No	g3033.t1
LH	Autosomes	GO:0008046	mf	axon guidance receptor	1	0.14478	0.2408	No	g3532.t1

				activity					
LH	Autosomes	GO:0016199	bp	axon midline choice point recognition	1	0.13868	0.2373	No	g3532.t1
LH	Autosomes	GO:0035385	bp	Roundabout signaling pathway	1	0.13868	0.2373	No	g3532.t1
LH	Autosomes	GO:0005634	cc	nucleus	5	0.07888	0.1844	No	g3792.t1,g6358.t1,g9044.t1, g14788.t1,g14822.t1
LH	Autosomes	GO:0000049	mf	tRNA binding	1	0.19209	0.2822	No	g3792.t1
LH	Autosomes	GO:0004672	mf	protein kinase activity	1	0.626	0.6974	No	g3792.t1
LH	Autosomes	GO:0005524	mf	ATP binding	6	0.48238	0.56	No	g3792.t1,g5456.t1,g9033.t1, g14814.t1,g14908.t1,g15089. t1
LH	Autosomes	GO:0031267	mf	small GTPase binding	2	0.19851	0.2878	No	g3792.t1,g5456.t1
LH	Autosomes	GO:0006468	bp	protein phosphorylation	3	0.52357	0.6057	No	g3792.t1,g5456.t1,g15089.t1
LH	Autosomes	GO:0006886	bp	intracellular protein transport	1	1	1	No	g3792.t1
LH	Autosomes	GO:0071528	bp	tRNA re-export from nucleus	1	0.00708	0.0712	No	g3792.t1
LH	Autosomes	GO:0022625	cc	cytosolic large ribosomal subunit	1	0.17467	0.2636	No	g3793.t1
LH	Autosomes	GO:0042788	cc	polysomal ribosome	1	0.08176	0.1872	No	g3793.t1
LH	Autosomes	GO:0003735	mf	structural constituent of ribosome	2	0.15803	0.2429	No	g3793.t1,g6359.t1
LH	Autosomes	GO:0002181	bp	cytoplasmic translation	1	0.14478	0.2408	No	g3793.t1
LH	Autosomes	GO:0005764	cc	lysosome	1	0.39235	0.4771	No	g3794.t1
LH	Autosomes	GO:0045171	cc	intercellular bridge	1	0.15686	0.2422	No	g3794.t1
LH	Autosomes	GO:0140007	cc	KICSTOR complex	1	0.02109	0.0986	No	g3794.t1
LH	Autosomes	GO:0034198	bp	cellular response to amino acid starvation	1	0.13253	0.2328	No	g3794.t1
LH	Autosomes	GO:0042149	bp	cellular response to glucose starvation	1	0.1075	0.2099	No	g3794.t1
LH	Autosomes	GO:0061462	bp	protein localization to	1	0.02803	0.1082	No	g3794.t1

				lysosome					
LH	Autosomes	GO:1904262	bp	negative regulation of TORC1 signaling	1	0.06861	0.1663	No	g3794.t1
LH	Autosomes	GO:0004719	mf	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity	1	0.02109	0.0986	No	g5303.t1
LH	Autosomes	GO:0006479	bp	protein methylation	1	0.05527	0.1468	No	g5303.t1
LH	Autosomes	GO:0016579	bp	protein deubiquitination	2	0.03723	0.1212	No	g5456.t1,g14877.t1
LH	Autosomes	GO:0070628	mf	proteasome binding	1	0.02803	0.1082	No	g5456.t1
LH	Autosomes	GO:0007268	bp	chemical synaptic transmission	1	0.24759	0.3369	No	g5456.t1
LH	Autosomes	GO:0000502	cc	proteasome complex	1	0.10113	0.2047	No	g5456.t1
LH	Autosomes	GO:0007266	bp	Rho protein signal transduction	1	0.09472	0.199	No	g5456.t1
LH	Autosomes	GO:0045202	cc	synapse	1	1	1	No	g5456.t1
LH	Autosomes	GO:0004674	mf	protein serine/threonine kinase activity	2	0.37245	0.4625	No	g5456.t1,g15089.t1
LH	Autosomes	GO:0005886	cc	plasma membrane	6	0.58629	0.6666	No	g5456.t1,g7045.t1,g7071.t1, g9044.t1,g14789.t1,g14908.t 1
LH	Autosomes	GO:0009986	cc	cell surface	2	0.19594	0.2853	No	g5456.t1,g14907.t1
LH	Autosomes	GO:0004843	mf	cysteine-type deubiquitinase activity	2	0.08387	0.1907	No	g5456.t1,g14877.t1
LH	Autosomes	GO:0005856	cc	cytoskeleton	3	0.18153	0.2703	No	g5456.t1,g16418.t1,g19406.t 1
LH	Autosomes	GO:0004866	mf	endopeptidase inhibitor activity	1	0.08826	0.1954	No	g5456.t1
LH	Autosomes	GO:0005829	cc	cytosol	8	0.69665	0.7659	No	g5456.t1,g9007.t1,g14875.t1 ,g14907.t1,g14909.t1,g14968 .t1,g16419.t1,g16421.t1

LH	Autosomes	GO:0031410	cc	cytoplasmic vesicle	2	0.05985	0.1565	No	g5456.t1,g14907.t1
LH	Autosomes	GO:0050920	bp	regulation of chemotaxis	1	0.02109	0.0986	No	g5456.t1
LH	Autosomes	GO:1903070	bp	negative regulation of ER-associated ubiquitin-dependent protein catabolic process	1	0.02109	0.0986	No	g5456.t1
LH	Autosomes	GO:0005615	cc	extracellular space	2	1	1	No	g6351.t1,g22291.t1
LH	Autosomes	GO:0019773	cc	proteasome core complex, alpha-subunit complex	1	0.03491	0.1159	No	g6358.t1
LH	Autosomes	GO:0006511	bp	ubiquitin-dependent protein catabolic process	2	0.13604	0.2373	No	g6358.t1,g16419.t1
LH	Autosomes	GO:0005762	cc	mitochondrial large ribosomal subunit	1	0.09472	0.199	No	g6359.t1
LH	Autosomes	GO:0006412	bp	translation	1	0.40095	0.4823	No	g6359.t1
LH	Autosomes	GO:0000166	mf	nucleotide binding	1	0.42603	0.5051	No	g6366.t1
LH	Autosomes	GO:0004074	mf	biliverdin reductase (NAD(P)+) activity	1	0.00708	0.0712	No	g6366.t1
LH	Autosomes	GO:0008270	mf	zinc ion binding	3	1	1	No	g6366.t1,g7039.t1,g14870.t1
LH	Autosomes	GO:0042167	bp	heme catabolic process	1	0.02109	0.0986	No	g6366.t1
LH	Autosomes	GO:0016791	mf	phosphatase activity	1	0.12011	0.2215	No	g6471.t1
LH	Autosomes	GO:0016311	bp	dephosphorylation	1	0.32381	0.4167	No	g6471.t1
LH	Autosomes	GO:0004222	mf	metalloendopeptidase activity	1	0.59829	0.6743	No	g7039.t1
LH	Autosomes	GO:0004930	mf	G protein-coupled receptor activity	1	0.72707	0.7888	No	g7045.t1
LH	Autosomes	GO:0007186	bp	G protein-coupled receptor signaling pathway	1	0.53159	0.6107	No	g7045.t1
LH	Autosomes	GO:0072659	bp	protein localization to	1	0.21477	0.3047	No	g7045.t1

				plasma membrane					
LH	Autosomes	GO:0008360	bp	regulation of cell shape	1	0.24759	0.3369	No	g7071.t1
LH	Autosomes	GO:0032092	bp	positive regulation of protein binding	1	0.15085	0.2408	No	g7071.t1
LH	Autosomes	GO:0008284	bp	positive regulation of cell population proliferation	2	0.13125	0.2328	No	g7071.t1,g16421.t1
LH	Autosomes	GO:0090263	bp	positive regulation of canonical Wnt signaling pathway	1	0.16877	0.2559	No	g7071.t1
LH	Autosomes	GO:0045668	bp	negative regulation of osteoblast differentiation	1	0.10113	0.2047	No	g7071.t1
LH	Autosomes	GO:0030514	bp	negative regulation of BMP signaling pathway	1	0.09472	0.199	No	g7071.t1
LH	Autosomes	GO:0010718	bp	positive regulation of epithelial to mesenchymal transition	2	0.00493	0.0712	No	g7071.t1,g16421.t1
LH	Autosomes	GO:0005109	mf	frizzled binding	1	0.10113	0.2047	No	g7071.t1
LH	Autosomes	GO:0045893	bp	positive regulation of DNA-templated transcription	2	0.19594	0.2853	No	g7071.t1,g14790.t1
LH	Autosomes	GO:0016477	bp	cell migration	4	0.00092	0.0614	No	g7071.t1,g9011.t1,g12447.t1 ,g12448.t1
LH	Autosomes	GO:0030512	bp	negative regulation of transforming growth factor beta receptor signaling pathway	1	0.15686	0.2422	No	g7071.t1
LH	Autosomes	GO:0005154	mf	epidermal growth factor receptor binding	2	0.00493	0.0712	No	g8934.t1,g8935.t1
LH	Autosomes	GO:0019901	mf	protein kinase binding	2	0.22961	0.3214	No	g8934.t1,g8935.t1
LH	Autosomes	GO:0036312	mf	phosphatidylinositol 3-	2	0.0003	0.0246	**Yes**	g8934.t1,g8935.t1

				kinase regulatory subunit binding					
LH	Autosomes	GO:0036313	mf	phosphatidylinositol 3-kinase catalytic subunit binding	2	4.98E-05	0.0157	**Yes**	g8934.t1,g8935.t1
LH	Autosomes	GO:0007173	bp	epidermal growth factor receptor signaling pathway	2	0.00314	0.0712	No	g8934.t1,g8935.t1
LH	Autosomes	GO:0008283	bp	cell population proliferation	2	0.02966	0.1132	No	g8934.t1,g8935.t1
LH	Autosomes	GO:0009898	cc	cytoplasmic side of plasma membrane	1	0.09472	0.199	No	g9007.t1
LH	Autosomes	GO:0042383	cc	sarcolemma	1	0.21477	0.3047	No	g9007.t1
LH	Autosomes	GO:0003993	mf	acid phosphatase activity	1	0.01411	0.0919	No	g9007.t1
LH	Autosomes	GO:0004726	mf	non-membrane spanning protein tyrosine phosphatase activity	1	0.02803	0.1082	No	g9007.t1
LH	Autosomes	GO:0035335	bp	peptidyl-tyrosine dephosphorylation	1	0.42192	0.5039	No	g9007.t1
LH	Autosomes	GO:0005576	cc	extracellular region	2	0.77629	0.8341	No	g9008.t1,g15653.t1
LH	Autosomes	GO:0030298	mf	receptor signaling protein tyrosine kinase activator activity	1	0.00708	0.0712	No	g9008.t1
LH	Autosomes	GO:0030971	mf	receptor tyrosine kinase binding	1	0.15085	0.2408	No	g9008.t1
LH	Autosomes	GO:0061098	bp	positive regulation of protein tyrosine kinase activity	1	0.09472	0.199	No	g9008.t1
LH	Autosomes	GO:0031965	cc	nuclear membrane	1	0.3092	0.401	No	g9011.t1
LH	Autosomes	GO:0003677	mf	DNA binding	3	0.80302	0.86	No	g9011.t1,g9033.t1,g14822.t1
LH	Autosomes	GO:0005789	cc	endoplasmic reticulum membrane	3	0.43761	0.517	No	g9033.t1,g14906.t1,g16419.t1

LH	Autosomes	GO:0102158	mf	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase activity	1	0.15085	0.2408	No	g9033.t1
LH	Autosomes	GO:0102343	mf	obsolete 3-hydroxy-arachidoyl-CoA dehydratase activity	1	0.15085	0.2408	No	g9033.t1
LH	Autosomes	GO:0102344	mf	obsolete 3-hydroxy-behenoyl-CoA dehydratase activity	1	0.15085	0.2408	No	g9033.t1
LH	Autosomes	GO:0102345	mf	obsolete 3-hydroxy-lignoceroyl-CoA dehydratase activity	1	0.15085	0.2408	No	g9033.t1
LH	Autosomes	GO:0006633	bp	fatty acid biosynthetic process	1	0.22586	0.3177	No	g9033.t1
LH	Autosomes	GO:0032508	bp	DNA duplex unwinding	1	0.3428	0.4377	No	g9033.t1
LH	Autosomes	GO:0000776	cc	kinetochore	1	0.20351	0.2925	No	g9043.t1
LH	Autosomes	GO:0005793	cc	endoplasmic reticulum-Golgi intermediate compartment	1	0.10113	0.2047	No	g9043.t1
LH	Autosomes	GO:0005794	cc	Golgi apparatus	1	1	1	No	g9043.t1
LH	Autosomes	GO:0030008	cc	TRAPP complex	1	0.02803	0.1082	No	g9043.t1
LH	Autosomes	GO:0048471	cc	perinuclear region of cytoplasm	1	1	1	No	g9043.t1
LH	Autosomes	GO:0006888	bp	endoplasmic reticulum to Golgi vesicle-mediated transport	1	0.24222	0.3337	No	g9043.t1
LH	Autosomes	GO:0007030	bp	Golgi organization	2	0.02545	0.1082	No	g9043.t1,g13136.t1
LH	Autosomes	GO:0051310	bp	metaphase plate congression	1	0.02109	0.0986	No	g9043.t1
LH	Autosomes	GO:0090234	bp	regulation of kinetochore assembly	1	0.01411	0.0919	No	g9043.t1
LH	Autosomes	GO:1905342	bp	positive regulation of protein localization to	1	0.01411	0.0919	No	g9043.t1

				kinetochore					
LH	Autosomes	GO:0000287	mf	magnesium ion binding	1	0.53661	0.6143	No	g9044.t1
LH	Autosomes	GO:0003676	mf	nucleic acid binding	2	0.72303	0.7888	No	g9044.t1,g15655.t1
LH	Autosomes	GO:0004523	mf	RNA-DNA hybrid ribonuclease activity	2	0.03883	0.1252	No	g9044.t1,g15655.t1
LH	Autosomes	GO:0005506	mf	iron ion binding	2	0.14087	0.2398	No	g9044.t1,g14906.t1
LH	Autosomes	GO:0010309	mf	acireductone dioxygenase [iron(II)-requiring] activity	1	0.00708	0.0712	No	g9044.t1
LH	Autosomes	GO:0019509	bp	L-methionine salvage from methylthioadenosine	1	0.03491	0.1159	No	g9044.t1
LH	Autosomes	GO:0090502	bp	RNA phosphodiester bond hydrolysis, endonucleolytic	2	0.07968	0.185	No	g9044.t1,g15655.t1
LH	Autosomes	GO:0016787	mf	hydrolase activity	1	1	1	No	g9071.t1
LH	Autosomes	GO:0016746	mf	acyltransferase activity	1	0.13868	0.2373	No	g9084.t1
LH	Autosomes	GO:0030036	bp	actin cytoskeleton organization	2	0.06175	0.157	No	g12447.t1,g12448.t1
LH	Autosomes	GO:0000139	cc	Golgi membrane	3	0.1506	0.2408	No	g13136.t1,g14869.t1,g15562.t1
LH	Autosomes	GO:0110165	cc	cellular anatomical entity	1	0.60115	0.6743	No	g13137.t1
LH	Autosomes	GO:0004197	mf	cysteine-type endopeptidase activity	1	0.15686	0.2422	No	g13137.t1
LH	Autosomes	GO:0051603	bp	proteolysis involved in protein catabolic process	1	0.06196	0.157	No	g13137.t1
LH	Autosomes	GO:0030246	mf	carbohydrate binding	1	0.44217	0.5206	No	g14286.t1
LH	Autosomes	GO:0006897	bp	endocytosis	1	0.47685	0.5555	No	g14286.t1
LH	Autosomes	GO:0035861	cc	site of double-strand break	1	0.12011	0.2215	No	g14788.t1

LH	Autosomes	GO:0045830	bp	positive regulation of isotype switching	1	0.02109	0.0986	No	g14788.t1
LH	Autosomes	GO:2000042	bp	negative regulation of double-strand break repair via homologous recombination	1	0.05527	0.1468	No	g14788.t1
LH	Autosomes	GO:2001034	bp	positive regulation of double-strand break repair via nonhomologous end joining	1	0.02109	0.0986	No	g14788.t1
LH	Autosomes	GO:0031175	bp	neuron projection development	1	0.12011	0.2215	No	g14789.t1
LH	Autosomes	GO:0044754	cc	autolysosome	1	0.03491	0.1159	No	g14790.t1
LH	Autosomes	GO:0003713	mf	transcription coactivator activity	2	0.0656	0.165	No	g14790.t1,g16421.t1
LH	Autosomes	GO:0006622	bp	protein targeting to lysosome	1	0.04853	0.1365	No	g14790.t1
LH	Autosomes	GO:0006879	bp	intracellular iron ion homeostasis	1	0.15686	0.2422	No	g14790.t1
LH	Autosomes	GO:0009725	bp	response to hormone	1	0.04853	0.1365	No	g14790.t1
LH	Autosomes	GO:0005744	cc	TIM23 mitochondrial import inner membrane translocase complex	1	0.02109	0.0986	No	g14791.t1
LH	Autosomes	GO:0005758	cc	mitochondrial intermembrane space	1	0.16877	0.2559	No	g14791.t1
LH	Autosomes	GO:0031305	cc	obsolete integral component of mitochondrial inner membrane	1	0.06861	0.1663	No	g14791.t1
LH	Autosomes	GO:0008320	mf	protein transmembrane transporter activity	1	0.03491	0.1159	No	g14791.t1

LH	Autosomes	GO:0030150	bp	protein import into mitochondrial matrix	1	0.04853	0.1365	No	g14791.t1
LH	Autosomes	GO:0004478	mf	methionine adenosyltransferase activity	1	0.01411	0.0919	No	g14814.t1
LH	Autosomes	GO:0046872	mf	metal ion binding	6	7.10E-01	0.7774	No	g14814.t1,g15655.t1,g16419.t1,g22289.t1,g22290.t1,g22292.t1
LH	Autosomes	GO:0006556	bp	S-adenosylmethionine biosynthetic process	1	0.02803	0.1082	No	g14814.t1
LH	Autosomes	GO:0006730	bp	one-carbon metabolic process	1	0.09472	0.199	No	g14814.t1
LH	Autosomes	GO:0005581	cc	collagen trimer	1	0.26871	0.3569	No	g14817.t1
LH	Autosomes	GO:0005730	cc	nucleolus	2	0.69321	0.7646	No	g14821.t1,g14906.t1
LH	Autosomes	GO:0005525	mf	GTP binding	2	1	1	No	g14821.t1,g19406.t1
LH	Autosomes	GO:0042254	bp	ribosome biogenesis	1	0.12011	0.2215	No	g14821.t1
LH	Autosomes	GO:0015016	mf	[heparan sulfate]-glucosamine N-sulfotransferase activity	1	0.03491	0.1159	No	g14869.t1
LH	Autosomes	GO:0019213	mf	deacetylase activity	1	0.02109	0.0986	No	g14869.t1
LH	Autosomes	GO:0002002	bp	regulation of angiotensin levels in blood	1	0.00708	0.0712	No	g14869.t1
LH	Autosomes	GO:0002448	bp	mast cell mediated immunity	1	0.01411	0.0919	No	g14869.t1
LH	Autosomes	GO:0015014	bp	heparan sulfate proteoglycan biosynthetic process, polysaccharide chain biosynthetic process	1	0.03491	0.1159	No	g14869.t1
LH	Autosomes	GO:0030210	bp	heparin biosynthetic process	1	0.05527	0.1468	No	g14869.t1
LH	Autosomes	GO:0015629	cc	actin cytoskeleton	2	0.04377	0.1309	No	g14875.t1,g14907.t1

LH	Autosomes	GO:0016607	cc	nuclear speck	1	0.59541	0.6743	No	g14875.t1
LH	Autosomes	GO:0030018	cc	Z disc	2	0.02411	0.1082	No	g14875.t1,g14876.t1
LH	Autosomes	GO:0030054	cc	cell junction	1	0.25293	0.34	No	g14875.t1
LH	Autosomes	GO:0003779	mf	actin binding	1	1	1	No	g14875.t1
LH	Autosomes	GO:0032233	bp	positive regulation of actin filament bundle assembly	1	0.02109	0.0986	No	g14875.t1
LH	Autosomes	GO:0004768	mf	stearoyl-CoA 9-desaturase activity	1	0.01411	0.0919	No	g14906.t1
LH	Autosomes	GO:0006636	bp	unsaturated fatty acid biosynthetic process	1	0.08826	0.1954	No	g14906.t1
LH	Autosomes	GO:0005783	cc	endoplasmic reticulum	2	0.65856	0.7288	No	g14907.t1,g15653.t1
LH	Autosomes	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	1	0.73045	0.7899	No	g14907.t1
LH	Autosomes	GO:0034704	cc	calcium channel complex	1	0.06861	0.1663	No	g14907.t1
LH	Autosomes	GO:0043235	cc	receptor complex	1	0.25293	0.34	No	g14907.t1
LH	Autosomes	GO:0060170	cc	ciliary membrane	1	0.11383	0.2159	No	g14907.t1
LH	Autosomes	GO:0097730	cc	non-motile cilium	1	0.04853	0.1365	No	g14907.t1
LH	Autosomes	GO:0005262	mf	calcium channel activity	1	0.20916	0.2993	No	g14907.t1
LH	Autosomes	GO:0005272	mf	sodium channel activity	1	0.09472	0.199	No	g14907.t1
LH	Autosomes	GO:0005509	mf	calcium ion binding	1	0.27115	0.3586	No	g14907.t1
LH	Autosomes	GO:0015269	mf	calcium-activated potassium channel activity	1	0.02803	0.1082	No	g14907.t1
LH	Autosomes	GO:0033040	mf	sour taste receptor activity	1	0.00708	0.0712	No	g14907.t1
LH	Autosomes	GO:0042802	mf	identical protein binding	2	1	1	No	g14907.t1,g19405.t1
LH	Autosomes	GO:0051371	mf	muscle alpha-actinin binding	1	0.01411	0.0919	No	g14907.t1
LH	Autosomes	GO:0001581	bp	detection of chemical stimulus involved in	1	0.00708	0.0712	No	g14907.t1

				sensory perception of sour taste					
LH	Autosomes	GO:0007224	bp	smoothened signaling pathway	1	0.13253	0.2328	No	g14907.t1
LH	Autosomes	GO:0009415	bp	response to water	1	0.00708	0.0712	No	g14907.t1
LH	Autosomes	GO:0035725	bp	sodium ion transmembrane transport	1	0.388	0.4736	No	g14907.t1
LH	Autosomes	GO:0051289	bp	protein homotetramerization	1	0.12634	0.2267	No	g14907.t1
LH	Autosomes	GO:0070588	bp	calcium ion transmembrane transport	1	0.3658	0.4566	No	g14907.t1
LH	Autosomes	GO:0071468	bp	cellular response to acidic pH	1	0.02803	0.1082	No	g14907.t1
LH	Autosomes	GO:0071805	bp	potassium ion transmembrane transport	1	0.61515	0.6876	No	g14907.t1
LH	Autosomes	GO:0031966	cc	mitochondrial membrane	1	0.2368	0.3276	No	g14908.t1
LH	Autosomes	GO:0008559	mf	ABC-type xenobiotic transporter activity	1	0.02109	0.0986	No	g14908.t1
LH	Autosomes	GO:0042908	bp	xenobiotic transport	1	0.06196	0.157	No	g14908.t1
LH	Autosomes	GO:0055085	bp	transmembrane transport	1	1	1	No	g14908.t1
LH	Autosomes	GO:0000930	cc	gamma-tubulin complex	1	0.01411	0.0919	No	g14909.t1
LH	Autosomes	GO:0005739	cc	mitochondrion	3	0.24946	0.338	No	g14909.t1,g15653.t1,g22290.t1
LH	Autosomes	GO:0031083	cc	BLOC-1 complex	1	0.03491	0.1159	No	g14909.t1
LH	Autosomes	GO:0099078	cc	BORC complex	1	0.00708	0.0712	No	g14909.t1
LH	Autosomes	GO:1904115	cc	axon cytoplasm	1	0.11383	0.2159	No	g14909.t1
LH	Autosomes	GO:0043015	mf	gamma-tubulin binding	1	0.11383	0.2159	No	g14909.t1

LH	Autosomes	GO:0008625	bp	extrinsic apoptotic signaling pathway via death domain receptors	1	0.03491	0.1159	No	g14909.t1
LH	Autosomes	GO:0016197	bp	endosomal transport	1	0.08176	0.1872	No	g14909.t1
LH	Autosomes	GO:0032418	bp	lysosome localization	1	0.04174	0.126	No	g14909.t1
LH	Autosomes	GO:0048490	bp	anterograde synaptic vesicle transport	1	0.04853	0.1365	No	g14909.t1
LH	Autosomes	GO:0097345	bp	mitochondrial outer membrane permeabilization	1	0.02109	0.0986	No	g14909.t1
LH	Autosomes	GO:0016788	mf	hydrolase activity, acting on ester bonds	3	9.47E-05	0.0157	**Yes**	g14938.t1,g14939.t1,g14941.t1
LH	Autosomes	GO:0016042	bp	lipid catabolic process	3	0.00598	0.0712	No	g14938.t1,g14939.t1,g14968.t1
LH	Autosomes	GO:0006629	bp	lipid metabolic process	1	0.3521	0.4462	No	g14941.t1
LH	Autosomes	GO:0004435	mf	phosphatidylinositol phospholipase C activity	1	0.12634	0.2267	No	g14968.t1
LH	Autosomes	GO:0005085	mf	guanyl-nucleotide exchange factor activity	1	1	1	No	g14968.t1
LH	Autosomes	GO:0007264	bp	small GTPase mediated signal transduction	1	0.3567	0.4486	No	g14968.t1
LH	Autosomes	GO:0050790	bp	regulation of catalytic activity	1	0.72673	0.7888	No	g14968.t1
LH	Autosomes	GO:0008467	mf	[heparan sulfate]-glucosamine 3-sulfotransferase 1 activity	1	0.02109	0.0986	No	g14991.t1
LH	Autosomes	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	1	1	1	No	g15089.t1
LH	Autosomes	GO:0106310	mf	protein serine kinase activity	1	0.60115	0.6743	No	g15089.t1

LH	Autosomes	GO:0005229	mf	intracellular calcium activated chloride channel activity	1	0.02803	0.1082	No	g15561.t1
LH	Autosomes	GO:0008237	mf	metallopeptidase activity	1	0.1075	0.2099	No	g15561.t1
LH	Autosomes	GO:0006821	bp	chloride transport	1	0.25823	0.3443	No	g15561.t1
LH	Autosomes	GO:0005741	cc	mitochondrial outer membrane	1	0.27905	0.3662	No	g15562.t1
LH	Autosomes	GO:0008289	mf	lipid binding	1	0.37477	0.4625	No	g15562.t1
LH	Autosomes	GO:0006915	bp	apoptotic process	1	0.42603	0.5051	No	g15562.t1
LH	Autosomes	GO:0061024	bp	membrane organization	1	0.01411	0.0919	No	g15562.t1
LH	Autosomes	GO:0030425	cc	dendrite	1	0.45397	0.5326	No	g15653.t1
LH	Autosomes	GO:0071300	bp	cellular response to retinoic acid	1	0.0752	0.1771	No	g15653.t1
LH	Autosomes	GO:0045930	bp	negative regulation of mitotic cell cycle	1	0.05527	0.1468	No	g15653.t1
LH	Autosomes	GO:0045666	bp	positive regulation of neuron differentiation	1	0.18633	0.2749	No	g15653.t1
LH	Autosomes	GO:0043025	cc	neuronal cell body	1	0.36127	0.4526	No	g15653.t1
LH	Autosomes	GO:0016887	mf	ATP hydrolysis activity	1	1	1	No	g16418.t1
LH	Autosomes	GO:0032781	bp	positive regulation of ATP-dependent activity	1	0.06196	0.157	No	g16418.t1
LH	Autosomes	GO:1900028	bp	negative regulation of ruffle assembly	1	0.02109	0.0986	No	g16418.t1
LH	Autosomes	GO:0003785	mf	actin monomer binding	1	0.08826	0.1954	No	g16418.t1
LH	Autosomes	GO:0098685	cc	Schaffer collateral - CA1 synapse	1	0.16284	0.2491	No	g16418.t1
LH	Autosomes	GO:0005546	mf	phosphatidylinositol-4,5-bisphosphate binding	1	0.12634	0.2267	No	g16418.t1
LH	Autosomes	GO:0098793	cc	presynapse	1	0.24759	0.3369	No	g16418.t1
LH	Autosomes	GO:0098794	cc	postsynapse	1	0.14478	0.2408	No	g16418.t1
LH	Autosomes	GO:0030837	bp	negative regulation of actin filament	1	0.04174	0.126	No	g16418.t1

				polymerization					
LH	Autosomes	GO:0051496	bp	positive regulation of stress fiber assembly	1	0.14478	0.2408	No	g16418.t1
LH	Autosomes	GO:0098885	bp	modification of postsynaptic actin cytoskeleton	1	0.00708	0.0712	No	g16418.t1
LH	Autosomes	GO:0030838	bp	positive regulation of actin filament polymerization	1	0.0752	0.1771	No	g16418.t1
LH	Autosomes	GO:0010633	bp	negative regulation of epithelial cell migration	1	0.02803	0.1082	No	g16418.t1
LH	Autosomes	GO:2000300	bp	regulation of synaptic vesicle exocytosis	1	0.1075	0.2099	No	g16418.t1
LH	Autosomes	GO:0098978	cc	glutamatergic synapse	1	0.46553	0.5442	No	g16418.t1
LH	Autosomes	GO:0050821	bp	protein stabilization	1	0.2739	0.3609	No	g16418.t1
LH	Autosomes	GO:0033138	bp	positive regulation of peptidyl-serine phosphorylation	1	0.15686	0.2422	No	g16418.t1
LH	Autosomes	GO:0005765	cc	lysosomal membrane	1	0.39666	0.4789	No	g16419.t1
LH	Autosomes	GO:0031902	cc	late endosome membrane	1	0.22586	0.3177	No	g16419.t1
LH	Autosomes	GO:0061630	mf	ubiquitin protein ligase activity	1	0.5592	0.638	No	g16419.t1
LH	Autosomes	GO:0051865	bp	protein autoubiquitination	1	0.15686	0.2422	No	g16419.t1
LH	Autosomes	GO:0070304	bp	positive regulation of stress-activated protein kinase signaling cascade	1	0.00708	0.0712	No	g16419.t1
LH	Autosomes	GO:0035264	bp	multicellular organism growth	1	0.18633	0.2749	No	g16421.t1
LH	Autosomes	GO:0016604	cc	nuclear body	1	0.38362	0.47	No	g16421.t1

LH	Autosomes	GO:0045669	bp	positive regulation of osteoblast differentiation	1	0.15686	0.2422	No	g16421.t1
LH	Autosomes	GO:0045944	bp	positive regulation of transcription by RNA polymerase II	2	1	1	No	g16421.t1,g22291.t1
LH	Autosomes	GO:0006469	bp	negative regulation of protein kinase activity	1	0.12634	0.2267	No	g16421.t1
LH	Autosomes	GO:0003015	bp	heart process	1	0.00708	0.0712	No	g16421.t1
LH	Autosomes	GO:0000122	bp	negative regulation of transcription by RNA polymerase II	1	1	1	No	g16421.t1
LH	Autosomes	GO:0005667	cc	transcription regulator complex	1	0.41362	0.4957	No	g16421.t1
LH	Autosomes	GO:0060390	bp	regulation of SMAD protein signal transduction	1	0.02109	0.0986	No	g16421.t1
LH	Autosomes	GO:0017145	bp	stem cell division	1	0.03491	0.1159	No	g16421.t1
LH	Autosomes	GO:0072307	bp	regulation of metanephric nephron tubule epithelial cell differentiation	1	0.00708	0.0712	No	g16421.t1
LH	Autosomes	GO:1900182	bp	positive regulation of protein localization to nucleus	1	0.05527	0.1468	No	g16421.t1
LH	Autosomes	GO:0042803	mf	protein homodimerization activity	1	1	1	No	g16421.t1
LH	Autosomes	GO:0060993	bp	kidney morphogenesis	1	0.04174	0.126	No	g16421.t1
LH	Autosomes	GO:0060271	bp	cilium assembly	1	0.37477	0.4625	No	g16421.t1
LH	Autosomes	GO:0045599	bp	negative regulation of fat cell differentiation	1	0.10113	0.2047	No	g16421.t1

LH	Autosomes	GO:0090090	bp	negative regulation of canonical Wnt signaling pathway	1	0.28925	0.3781	No	g16421.t1
LH	Autosomes	GO:0001894	bp	tissue homeostasis	1	0.04174	0.126	No	g16421.t1
LH	Autosomes	GO:0032835	bp	glomerulus development	1	0.01411	0.0919	No	g16421.t1
LH	Autosomes	GO:0003714	mf	transcription corepressor activity	1	0.3567	0.4486	No	g16421.t1
LH	Autosomes	GO:0031146	bp	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	1	0.12011	0.2215	No	g16421.t1
LH	Autosomes	GO:0035329	bp	hippo signaling	1	0.04853	0.1365	No	g16421.t1
LH	Autosomes	GO:0016567	bp	protein ubiquitination	1	1	1	No	g16421.t1
LH	Autosomes	GO:1990779	cc	glycoprotein Ib-IX-V complex	1	0.00708	0.0712	No	g19405.t1
LH	Autosomes	GO:0007597	bp	blood coagulation, intrinsic pathway	1	0.00708	0.0712	No	g19405.t1
LH	Autosomes	GO:0010572	bp	positive regulation of platelet activation	1	0.01411	0.0919	No	g19405.t1
LH	Autosomes	GO:0051209	bp	release of sequestered calcium ion into cytosol	1	0.10113	0.2047	No	g19405.t1
LH	Autosomes	GO:0008021	cc	synaptic vesicle	1	0.23135	0.3214	No	g19406.t1
LH	Autosomes	GO:0017157	bp	regulation of exocytosis	1	0.11383	0.2159	No	g19406.t1
LH	Autosomes	GO:0003874	mf	6-pyruvoyltetrahydropterin synthase activity	1	0.00708	0.0712	No	g22289.t1
LH	Autosomes	GO:0006729	bp	tetrahydrobiopterin biosynthetic process	1	0.04174	0.126	No	g22289.t1
LH	Autosomes	GO:0003834	mf	beta-carotene 15,15'-dioxygenase activity	1	0.01411	0.0919	No	g22290.t1
LH	Autosomes	GO:0010436	mf	carotenoid dioxygenase activity	1	0.01411	0.0919	No	g22290.t1

LH	Autosomes	GO:0102076	mf	beta,beta-carotene-9',10'-cleaving oxygenase activity	1	0.00708	0.0712	No	g22290.t1
LH	Autosomes	GO:0016121	bp	carotene catabolic process	1	0.01411	0.0919	No	g22290.t1
LH	Autosomes	GO:0016124	bp	xanthophyll catabolic process	1	0.00708	0.0712	No	g22290.t1
LH	Autosomes	GO:0042574	bp	retinal metabolic process	1	0.05527	0.1468	No	g22290.t1
LH	Autosomes	GO:0051881	bp	regulation of mitochondrial membrane potential	1	0.0752	0.1771	No	g22290.t1
LH	Autosomes	GO:2000377	bp	regulation of reactive oxygen species metabolic process	1	0.03491	0.1159	No	g22290.t1
LH	Autosomes	GO:0042119	bp	neutrophil activation	1	0.02803	0.1082	No	g22291.t1
LH	Autosomes	GO:0014068	bp	positive regulation of phosphatidylinositol 3-kinase signaling	1	0.1075	0.2099	No	g22291.t1
LH	Autosomes	GO:1901224	bp	positive regulation of NIK/NF-kappaB signaling	1	0.1075	0.2099	No	g22291.t1
LH	Autosomes	GO:0042531	bp	positive regulation of tyrosine phosphorylation of STAT protein	1	0.03491	0.1159	No	g22291.t1
LH	Autosomes	GO:0042632	bp	cholesterol homeostasis	1	0.15085	0.2408	No	g22291.t1
LH	Autosomes	GO:0051092	bp	positive regulation of NF-kappaB transcription factor activity	1	0.20351	0.2925	No	g22291.t1
LH	Autosomes	GO:0048661	bp	positive regulation of smooth muscle cell proliferation	1	0.04174	0.126	No	g22291.t1
LH	Autosomes	GO:0000165	bp	MAPK cascade	1	0.13253	0.2328	No	g22291.t1
LH	Autosomes	GO:0006954	bp	inflammatory response	1	0.37921	0.4663	No	g22291.t1

LH	Autosomes	GO:0032725	bp	positive regulation of granulocyte macrophage colony-stimulating factor production	1	0.02109	0.0986	No	g22291.t1
LH	Autosomes	GO:0001525	bp	angiogenesis	1	0.30427	0.3961	No	g22291.t1
LH	Autosomes	GO:0032729	bp	positive regulation of type II interferon production	1	0.06861	0.1663	No	g22291.t1
LH	Autosomes	GO:0051142	bp	positive regulation of NK T cell proliferation	1	0.00708	0.0712	No	g22291.t1
LH	Autosomes	GO:0032740	bp	positive regulation of interleukin-17 production	1	0.02803	0.1082	No	g22291.t1
LH	Autosomes	GO:2000556	bp	positive regulation of T-helper 1 cell cytokine production	1	0.00708	0.0712	No	g22291.t1
LH	Autosomes	GO:0120162	bp	positive regulation of cold-induced thermogenesis	1	0.18052	0.27	No	g22291.t1
LH	Autosomes	GO:0070328	bp	triglyceride homeostasis	1	0.04853	0.1365	No	g22291.t1
LH	Autosomes	GO:0032148	bp	activation of protein kinase B activity	1	0.02803	0.1082	No	g22291.t1
LH	Autosomes	GO:0035655	bp	interleukin-18-mediated signaling pathway	1	0.01411	0.0919	No	g22291.t1
LH	Autosomes	GO:0045630	bp	positive regulation of T-helper 2 cell differentiation	1	0.01411	0.0919	No	g22291.t1
LH	Autosomes	GO:0042267	bp	natural killer cell mediated cytotoxicity	1	0.04174	0.126	No	g22291.t1
LH	Autosomes	GO:0005125	mf	cytokine activity	1	0.31897	0.4121	No	g22291.t1
LH	Autosomes	GO:0045515	mf	interleukin-18 receptor binding	1	0.00708	0.0712	No	g22291.t1

LH	Autosomes	GO:0042104	bp	positive regulation of activated T cell proliferation	1	0.02803	0.1082	No	g22291.t1
LH	Autosomes	GO:0032819	bp	positive regulation of natural killer cell proliferation	1	0.00708	0.0712	No	g22291.t1
LH	Autosomes	GO:0045662	bp	negative regulation of myoblast differentiation	1	0.05527	0.1468	No	g22291.t1
LH	Autosomes	GO:0031663	bp	lipopolysaccharide-mediated signaling pathway	1	0.0752	0.1771	No	g22291.t1
LH	Autosomes	GO:0010744	bp	positive regulation of macrophage derived foam cell differentiation	1	0.02109	0.0986	No	g22291.t1
LH	Autosomes	GO:0071407	bp	cellular response to organic cyclic compound	1	0.06861	0.1663	No	g22291.t1
LH	Autosomes	GO:0051897	bp	positive regulation of protein kinase B signaling	1	0.13868	0.2373	No	g22291.t1
LH	Autosomes	GO:0005743	cc	mitochondrial inner membrane	1	0.39666	0.4789	No	g22292.t1
LH	Autosomes	GO:0006099	bp	tricarboxylic acid cycle	1	0.15085	0.2408	No	g22292.t1
LH	Z	GO:0003964	mf	RNA-directed DNA polymerase activity	2	0.67865	0.6787	No	g10995.t1,g11417.t1
LH	Z	GO:0006278	bp	RNA-templated DNA biosynthetic process	2	0.67874	0.6787	No	g10995.t1,g11417.t1
LH	Z	GO:0016514	cc	SWI/SNF complex	1	0.01147	0.0281	**Yes**	g10997.t1
LH	Z	GO:0005524	mf	ATP binding	1	0.57689	0.6271	No	g10997.t1
LH	Z	GO:0016887	mf	ATP hydrolysis activity	1	0.15854	0.2154	No	g10997.t1
LH	Z	GO:0042393	mf	histone binding	1	0.02548	0.0425	**Yes**	g10997.t1
LH	Z	GO:0140658	mf	ATP-dependent chromatin remodeler activity	1	0.02349	0.0425	**Yes**	g10997.t1

LH	Z	GO:0006338	bp	chromatin remodeling	1	0.04581	0.0674	No	g10997.t1
LH	Z	GO:0006355	bp	regulation of DNA-templated transcription	1	0.1875	0.2344	No	g10997.t1
LH	Z	GO:0051276	bp	chromosome organization	1	0.00406	0.0203	**Yes**	g10997.t1
LH	Z	GO:0005905	cc	clathrin-coated pit	1	0.01349	0.0281	**Yes**	g11033.t1
LH	Z	GO:0072583	bp	clathrin-dependent endocytosis	1	0.00676	0.0242	**Yes**	g11033.t1
LH	Z	GO:0005778	cc	peroxisomal membrane	1	0.00609	0.0242	**Yes**	g11056.t1
LH	Z	GO:0005789	cc	endoplasmic reticulum membrane	1	0.16372	0.2154	No	g11056.t1
LH	Z	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	1	0.49369	0.561	No	g11056.t1
LH	Z	GO:0004420	mf	hydroxymethylglutaryl-CoA reductase (NADPH) activity	1	0.00068	0.0085	**Yes**	g11056.t1
LH	Z	GO:0050661	mf	NADP binding	1	0.01349	0.0281	**Yes**	g11056.t1
LH	Z	GO:0006695	bp	cholesterol biosynthetic process	1	0.0108	0.0281	**Yes**	g11056.t1
LH	Z	GO:0008299	bp	isoprenoid biosynthetic process	1	0.00406	0.0203	**Yes**	g11056.t1
LH	Z	GO:0015936	bp	coenzyme A metabolic process	1	0.00068	0.0085	**Yes**	g11056.t1
LH	Z	GO:0016592	cc	mediator complex	1	0.02548	0.0425	**Yes**	g11057.t1
LH	Z	GO:0070847	cc	core mediator complex	1	0.00811	0.0254	**Yes**	g11057.t1
LH	Z	GO:0003712	mf	transcription coregulator activity	1	0.03405	0.0532	No	g11057.t1
LH	Z	GO:0006357	bp	regulation of transcription by RNA polymerase II	1	0.19805	0.2358	No	g11057.t1
LH	Z	GO:0006369	bp	termination of RNA polymerase II	1	0.00203	0.0169	**Yes**	g11057.t1


transcription

Table S1.8: GO Term enrichment analyses for 196 Autosome genes and 21 Z-linked genes under MH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be FDR- adjusted p values.

Categories:	DNA Information Processing	Cellular structure, metabolism and function	Behaviour	Individual's well-being and health (e.g.sleep,appetite, mood, energy)	Neurotransmission	Fertilization & embryonic development	Immune response

Windows Type	Chromosome Type	GO ID	Ontology	GO Name	Number of Genes	p-value	Adjusted p-value	Significance	Genes
MH	Autosomes	GO:0003964	mf	RNA-directed DNA polymerase activity	42	4.81E-06	0.00074	**Yes**	g283.t1,g285.t1,g295.t1,g301.t1,g314.t1,g337.t1,g344.t1,g1711.t1,g1715.t1,g1720.t1,g3715.t1,g4692.t1,g4695.t1,g4696.t1,g4697.t1,g4705.t1,g4706.t1,g4708.t1,g4710.t1,g5440.t1,g5495.t1,g5589.t1,g5590.t1,g5591.t1,g5594.t1,g5595.t1,g5599.t1,g5601.t1,g5971.t1,g5972.t1,g6011.t1,g9122.t1,g9132.t1,g9152.t1,g13216.t1,g13230.t1,g13232.t1,g13233.t1,g16027.t1,g19203.t1,g19204.t1,g20997.t1
MH	Autosomes	GO:0006278	bp	RNA-templated DNA biosynthetic process	42	4.84E-06	0.00074	**Yes**	g283.t1,g285.t1,g295.t1,g301.t1,g314.t1,g337.t1,g344.t1,g1711.t1,g1715.t1,g1720.t1,g3715.t1,g4692.t1,g4695.t1,g4696.t1,g4697.t1,g4705.t1,g470

									6.t1,g4708.t1,g4710.t1,g5440.t1,g5495.t1,g5589.t1,g5590.t1,g5591.t1,g5594.t1,g5595.t1,g5599.t1,g5601.t1,g5971.t1,g5972.t1,g6011.t1,g9122.t1,g9132.t1,g9152.t1,g13216.t1,g13230.t1,g13232.t1,g13233.t1,g16027.t1,g19203.t1,g19204.t1,g20997.t1
MH	Autosomes	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	14	0.0005	0.03773	**Yes**	g295.t1,g335.t1,g362.t1,g1717.t1,g1719.t1,g5601.t1,g9132.t1,g16165.t1,g17966.t1,g20983.t1,g20990.t1,g22003.t1,g22358.t1,g22362.t1
MH	Autosomes	GO:0003824	mf	catalytic activity	6	0.3145	0.40512	No	g302.t1,g1742.t1,g3713.t1,g4712.t1,g5592.t1,g6603.t1
MH	Autosomes	GO:0005737	cc	cytoplasm	12	1	1	No	g318.t1,g333.t1,g4692.t1,g5434.t1,g5972.t1,g7253.t1,g17884.t1,g18481.t1,g19198.t1,g19207.t1,g20993.t1,g20996.t1
MH	Autosomes	GO:0005096	mf	GTPase activator activity	1	1	1	No	g318.t1
MH	Autosomes	GO:0007165	bp	signal transduction	3	1	1	No	g318.t1,g16025.t1,g20983.t1
MH	Autosomes	GO:0050790	bp	regulation of catalytic activity	1	0.73425	0.79435	No	g318.t1
MH	Autosomes	GO:0005634	cc	nucleus	5	0.02589	0.11573	No	g333.t1,g6012.t1,g9270.t1,g19205.t1,g22359.t1
MH	Autosomes	GO:0061630	mf	ubiquitin protein ligase activity	2	0.24783	0.34719	No	g333.t1,g19198.t1
MH	Autosomes	GO:0046872	mf	metal ion binding	4	0.11874	0.21109	No	g333.t1,g15252.t1,g19198.t1,g20981.t1

MH	Autosomes	GO:0043154	bp	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	1	0.11769	0.21076	No	g333.t1
MH	Autosomes	GO:0031398	bp	positive regulation of protein ubiquitination	1	0.14668	0.23845	No	g333.t1
MH	Autosomes	GO:0051726	bp	regulation of cell cycle	2	0.10976	0.20956	No	g333.t1,g20993.t1
MH	Autosomes	GO:0043027	mf	cysteine-type endopeptidase inhibitor activity involved in apoptotic process	1	0.04087	0.13218	No	g333.t1
MH	Autosomes	GO:0060546	bp	negative regulation of necroptotic process	1	0.04884	0.14701	No	g333.t1
MH	Autosomes	GO:0031012	cc	extracellular matrix	1	0.36841	0.44978	No	g336.t1
MH	Autosomes	GO:0004222	mf	metalloendopeptidase activity	1	1	1	No	g336.t1
MH	Autosomes	GO:0008270	mf	zinc ion binding	2	0.3359	0.42371	No	g336.t1,g19205.t1
MH	Autosomes	GO:0006508	bp	proteolysis	7	0.03552	0.13012	No	g336.t1,g4703.t1,g8745.t1,g13217.t1,g16165.t1,g17881.t1,g21003.t1
MH	Autosomes	GO:0035189	cc	Rb-E2F complex	1	0.01655	0.10706	No	g598.t1
MH	Autosomes	GO:0030308	bp	negative regulation of cell growth	1	0.19514	0.28976	No	g598.t1
MH	Autosomes	GO:0031175	bp	neuron projection development	1	0.13952	0.23051	No	g598.t1
MH	Autosomes	GO:0000977	mf	RNA polymerase II transcription regulatory region sequence-specific DNA binding	1	0.35775	0.44212	No	g598.t1
MH	Autosomes	GO:2000134	bp	negative regulation of G1/S transition of mitotic cell cycle	1	0.08008	0.17769	No	g598.t1

MH	Autosomes	GO:0048667	bp	cell morphogenesis involved in neuron differentiation	1	0.05675	0.15403	No	g598.t1
MH	Autosomes	GO:0006325	bp	chromatin organization	2	0.16324	0.2558	No	g598.t1,g19205.t1
MH	Autosomes	GO:0000785	cc	chromatin	1	0.54433	0.61745	No	g598.t1
MH	Autosomes	GO:0006357	bp	regulation of transcription by RNA polymerase II	2	1	1	No	g598.t1,g19199.t1
MH	Autosomes	GO:0016301	mf	kinase activity	4	0.12269	0.21474	No	g1712.t1,g1724.t1,g5971.t1,g16027.t1
MH	Autosomes	GO:0016310	bp	phosphorylation	5	0.0791	0.17769	No	g1712.t1,g1724.t1,g5971.t1,g16027.t1,g22004.t1
MH	Autosomes	GO:0110165	cc	cellular anatomical entity	1	1	1	No	g1713.t1
MH	Autosomes	GO:0009897	cc	external side of plasma membrane	1	0.3247	0.4165	No	g1716.t1
MH	Autosomes	GO:0043235	cc	receptor complex	1	0.28996	0.39004	No	g1716.t1
MH	Autosomes	GO:0004896	mf	cytokine receptor activity	1	0.1323	0.21979	No	g1716.t1
MH	Autosomes	GO:0019955	mf	cytokine binding	1	0.08773	0.18914	No	g1716.t1
MH	Autosomes	GO:0019221	bp	cytokine-mediated signaling pathway	1	0.24718	0.34719	No	g1716.t1
MH	Autosomes	GO:0005743	cc	mitochondrial inner membrane	2	0.11786	0.21076	No	g1717.t1,g5588.t1
MH	Autosomes	GO:0005471	mf	ATP:ADP antiporter activity	1	0.02473	0.11389	No	g1717.t1
MH	Autosomes	GO:0140021	bp	mitochondrial ADP transmembrane transport	1	0.02473	0.11389	No	g1717.t1
MH	Autosomes	GO:1901029	bp	negative regulation of mitochondrial outer membrane permeabilization involved in apoptotic signaling pathway	1	0.04087	0.13218	No	g1717.t1
MH	Autosomes	GO:1990544	bp	mitochondrial ATP transmembrane transport	1	0.02473	0.11389	No	g1717.t1

MH	Autosomes	GO:0015057	mf	thrombin-activated receptor activity	1	0.04884	0.14701	No	g1719.t1
MH	Autosomes	GO:0045028	mf	G protein-coupled purinergic nucleotide receptor activity	1	0.06459	0.1664	No	g1719.t1
MH	Autosomes	GO:0007596	bp	blood coagulation	1	0.18839	0.28212	No	g1719.t1
MH	Autosomes	GO:0035589	bp	G protein-coupled purinergic nucleotide receptor signaling pathway	1	0.07237	0.17187	No	g1719.t1
MH	Autosomes	GO:0070493	bp	thrombin-activated receptor signaling pathway	1	0.05675	0.15403	No	g1719.t1
MH	Autosomes	GO:0003723	mf	RNA binding	4	0.55583	0.62815	No	g1724.t1,g4692.t1,g5972.t1,g20993.t1
MH	Autosomes	GO:0051018	mf	protein kinase A binding	1	0.08008	0.17769	No	g1724.t1
MH	Autosomes	GO:0043484	bp	regulation of RNA splicing	1	0.11029	0.20956	No	g1724.t1
MH	Autosomes	GO:0005794	cc	Golgi apparatus	1	1	1	No	g4539.t1
MH	Autosomes	GO:0030414	mf	peptidase inhibitor activity	1	0.06459	0.1664	No	g4694.t1
MH	Autosomes	GO:0003676	mf	nucleic acid binding	3	0.50164	0.58205	No	g4699.t1,g4702.t1,g22358.t1
MH	Autosomes	GO:0004523	mf	RNA-DNA hybrid ribonuclease activity	2	0.05176	0.15277	No	g4699.t1,g4702.t1
MH	Autosomes	GO:0090502	bp	RNA phosphodiester bond hydrolysis, endonucleolytic	2	0.10444	0.20484	No	g4699.t1,g4702.t1
MH	Autosomes	GO:0015074	bp	DNA integration	1	0.26582	0.36732	No	g4702.t1
MH	Autosomes	GO:0004190	mf	aspartic-type endopeptidase activity	3	0.00357	0.09717	No	g4703.t1,g13217.t1,g21003.t1
MH	Autosomes	GO:0016772	mf	transferase activity, transferring phosphorus-containing groups	1	0.56303	0.63393	No	g4704.t1
MH	Autosomes	GO:0044237	bp	cellular metabolic process	1	0.57744	0.64066	No	g4704.t1
MH	Autosomes	GO:0005615	cc	extracellular space	1	0.26961	0.37071	No	g4705.t1

MH	Autosomes	GO:0004867	mf	serine-type endopeptidase inhibitor activity	1	0.59478	0.65276	No	g4705.t1
MH	Autosomes	GO:0010951	bp	negative regulation of endopeptidase activity	1	1	1	No	g4705.t1
MH	Autosomes	GO:0000940	cc	outer kinetochore	1	0.01655	0.10706	No	g5434.t1
MH	Autosomes	GO:0005654	cc	nucleoplasm	3	0.38542	0.4668	No	g5434.t1,g15214.t1,g19200.t1
MH	Autosomes	GO:0005813	cc	centrosome	2	0.30928	0.40512	No	g5434.t1,g15214.t1
MH	Autosomes	GO:0031262	cc	Ndc80 complex	1	0.02473	0.11389	No	g5434.t1
MH	Autosomes	GO:0030332	mf	cyclin binding	1	0.07237	0.17187	No	g5434.t1
MH	Autosomes	GO:0042802	mf	identical protein binding	2	1	1	No	g5434.t1,g17950.t1
MH	Autosomes	GO:0140483	mf	kinetochore adaptor activity	1	0.00831	0.09717	No	g5434.t1
MH	Autosomes	GO:0000132	bp	establishment of mitotic spindle orientation	1	0.05675	0.15403	No	g5434.t1
MH	Autosomes	GO:0007052	bp	mitotic spindle organization	1	0.11769	0.21076	No	g5434.t1
MH	Autosomes	GO:0007057	bp	spindle assembly involved in female meiosis I	1	0.00831	0.09717	No	g5434.t1
MH	Autosomes	GO:0008315	bp	G2/MI transition of meiotic cell cycle	1	0.00831	0.09717	No	g5434.t1
MH	Autosomes	GO:0031647	bp	regulation of protein stability	1	0.18839	0.28212	No	g5434.t1
MH	Autosomes	GO:0051298	bp	centrosome duplication	1	0.02473	0.11389	No	g5434.t1
MH	Autosomes	GO:0051301	bp	cell division	1	0.58793	0.64758	No	g5434.t1
MH	Autosomes	GO:0051315	bp	attachment of mitotic spindle microtubules to kinetochore	1	0.04884	0.14701	No	g5434.t1
MH	Autosomes	GO:0051383	bp	kinetochore organization	1	0.02473	0.11389	No	g5434.t1
MH	Autosomes	GO:0090267	bp	positive regulation of mitotic cell cycle spindle assembly checkpoint	1	0.01655	0.10706	No	g5434.t1

MH	Autosomes	GO:1905342	bp	positive regulation of protein localization to kinetochore	1	0.01655	0.10706	No	g5434.t1
MH	Autosomes	GO:0008046	mf	axon guidance receptor activity	1	0.16779	0.26025	No	g5496.t1
MH	Autosomes	GO:0016199	bp	axon midline choice point recognition	1	0.16081	0.2533	No	g5496.t1
MH	Autosomes	GO:0035385	bp	Roundabout signaling pathway	1	0.16081	0.2533	No	g5496.t1
MH	Autosomes	GO:0005488	mf	binding	1	0.57744	0.64066	No	g5585.t1
MH	Autosomes	GO:0140640	mf	catalytic activity, acting on a nucleic acid	1	0.12503	0.21474	No	g5585.t1
MH	Autosomes	GO:0044260	bp	obsolete cellular macromolecule metabolic process	1	0.15377	0.24734	No	g5585.t1
MH	Autosomes	GO:0090304	bp	nucleic acid metabolic process	1	0.24087	0.34333	No	g5585.t1
MH	Autosomes	GO:0034654	bp	nucleobase-containing compound biosynthetic process	1	0.01655	0.10706	No	g5588.t1
MH	Autosomes	GO:0016740	mf	transferase activity	2	0.24168	0.34333	No	g5593.t1,g7253.t1
MH	Autosomes	GO:0005085	mf	guanyl-nucleotide exchange factor activity	1	1	1	No	g5594.t1
MH	Autosomes	GO:0019208	mf	phosphatase regulator activity	1	0.10283	0.203	No	g5594.t1
MH	Autosomes	GO:0043087	bp	regulation of GTPase activity	1	0.3359	0.42371	No	g5594.t1
MH	Autosomes	GO:0008076	cc	voltage-gated potassium channel complex	4	0.00211	0.09717	No	g5596.t1,g21000.t1,g21001.t1,g21002.t1
MH	Autosomes	GO:0005249	mf	voltage-gated potassium channel activity	3	0.02217	0.11389	No	g5596.t1,g21000.t1,g21002.t1
MH	Autosomes	GO:0034765	bp	regulation of monoatomic ion transmembrane	5	0.00309	0.09717	No	g5596.t1,g14305.t1,g21000.t1,g21001.t1,g21002.t1

				transport					
MH	Autosomes	GO:0071805	bp	potassium ion transmembrane transport	4	0.02576	0.11573	No	g5596.t1,g21000.t1,g21001.t1,g21002.t1
MH	Autosomes	GO:0007049	bp	cell cycle	1	0.49191	0.57296	No	g6012.t1
MH	Autosomes	GO:0010468	bp	regulation of gene expression	1	0.21506	0.31584	No	g6012.t1
MH	Autosomes	GO:0005215	mf	transporter activity	1	0.03283	0.12172	No	g6545.t1
MH	Autosomes	GO:0016787	mf	hydrolase activity	2	0.35777	0.44212	No	g6545.t1,g20999.t1
MH	Autosomes	GO:0097159	mf	organic cyclic compound binding	1	0.1323	0.21979	No	g6545.t1
MH	Autosomes	GO:1901363	mf	heterocyclic compound binding	1	0.11769	0.21076	No	g6545.t1
MH	Autosomes	GO:0006812	bp	monoatomic cation transport	1	0.09531	0.19577	No	g6545.t1
MH	Autosomes	GO:0009987	bp	cellular process	1	1	1	No	g6545.t1
MH	Autosomes	GO:0065007	bp	biological regulation	1	0.11769	0.21076	No	g6545.t1
MH	Autosomes	GO:0004820	mf	glycine-tRNA ligase activity	1	0.01655	0.10706	No	g7253.t1
MH	Autosomes	GO:0005524	mf	ATP binding	4	0.05044	0.15033	No	g7253.t1,g7850.t1,g18480.t1,g20994.t1
MH	Autosomes	GO:0006426	bp	glycyl-tRNA aminoacylation	1	0.01655	0.10706	No	g7253.t1
MH	Autosomes	GO:1905653	bp	positive regulation of artery morphogenesis	1	0.00831	0.09717	No	g7850.t1
MH	Autosomes	GO:1905564	bp	positive regulation of vascular endothelial cell proliferation	1	0.04884	0.14701	No	g7850.t1
MH	Autosomes	GO:0048854	bp	brain morphogenesis	1	0.08773	0.18914	No	g7850.t1
MH	Autosomes	GO:0090050	bp	positive regulation of cell migration involved in sprouting angiogenesis	1	0.04087	0.13218	No	g7850.t1
MH	Autosomes	GO:0004674	mf	protein serine/threonine kinase activity	3	0.1954	0.28976	No	g7850.t1,g18480.t1,g20994.t1

MH	Autosomes	GO:0045766	bp	positive regulation of angiogenesis	1	0.18839	0.28212	No	g7850.t1
MH	Autosomes	GO:0000002	bp	mitochondrial genome maintenance	1	0.04087	0.13218	No	g7850.t1
MH	Autosomes	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	1	1	1	No	g7850.t1
MH	Autosomes	GO:0045793	bp	positive regulation of cell size	1	0.01655	0.10706	No	g7850.t1
MH	Autosomes	GO:0048873	bp	homeostasis of number of cells within a tissue	1	0.06459	0.1664	No	g7850.t1
MH	Autosomes	GO:0018105	bp	peptidyl-serine phosphorylation	2	0.06042	0.16253	No	g7850.t1,g18480.t1
MH	Autosomes	GO:0032008	bp	positive regulation of TOR signaling	1	0.11769	0.21076	No	g7850.t1
MH	Autosomes	GO:0106310	mf	protein serine kinase activity	1	1	1	No	g7850.t1
MH	Autosomes	GO:2000773	bp	negative regulation of cellular senescence	1	0.04087	0.13218	No	g7850.t1
MH	Autosomes	GO:0004252	mf	serine-type endopeptidase activity	2	0.14872	0.24049	No	g8745.t1,g16165.t1
MH	Autosomes	GO:0000276	cc	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	1	0.12503	0.21474	No	g9121.t1
MH	Autosomes	GO:0015078	mf	proton transmembrane transporter activity	1	0.1323	0.21979	No	g9121.t1
MH	Autosomes	GO:0006811	bp	monoatomic ion transport	1	0.27799	0.37728	No	g9121.t1
MH	Autosomes	GO:0015986	bp	proton motive force-driven ATP synthesis	1	0.14668	0.23845	No	g9121.t1
MH	Autosomes	GO:0002729	bp	positive regulation of natural killer cell cytokine production	1	0.02473	0.11389	No	g9132.t1

MH	Autosomes	GO:0003677	mf	DNA binding	1	0.06753	0.17187	No	g9270.t1
MH	Autosomes	GO:0000398	bp	mRNA splicing, via spliceosome	1	0.33033	0.42016	No	g9270.t1
MH	Autosomes	GO:0005840	cc	ribosome	1	0.33033	0.42016	No	g9271.t1
MH	Autosomes	GO:0003735	mf	structural constituent of ribosome	1	0.56668	0.63568	No	g9271.t1
MH	Autosomes	GO:0006412	bp	translation	1	0.45215	0.53277	No	g9271.t1
MH	Autosomes	GO:0001518	cc	voltage-gated sodium channel complex	1	0.09531	0.19577	No	g14305.t1
MH	Autosomes	GO:0005244	mf	voltage-gated monoatomic ion channel activity	1	0.20848	0.30766	No	g14305.t1
MH	Autosomes	GO:0005248	mf	voltage-gated sodium channel activity	1	0.11029	0.20956	No	g14305.t1
MH	Autosomes	GO:0035725	bp	sodium ion transmembrane transport	1	0.43822	0.52243	No	g14305.t1
MH	Autosomes	GO:0005814	cc	centriole	1	0.31332	0.40512	No	g15214.t1
MH	Autosomes	GO:0005829	cc	cytosol	5	0.36782	0.44978	No	g15214.t1,g19200.t1,g19206.t1,g22004.t1,g22359.t1
MH	Autosomes	GO:0019908	cc	nuclear cyclin-dependent protein kinase holoenzyme complex	1	0.00831	0.09717	No	g15214.t1
MH	Autosomes	GO:0097431	cc	mitotic spindle pole	1	0.05675	0.15403	No	g15214.t1
MH	Autosomes	GO:0019207	mf	kinase regulator activity	1	0.00831	0.09717	No	g15214.t1
MH	Autosomes	GO:0000079	bp	regulation of cyclin-dependent protein serine/threonine kinase activity	1	0.09531	0.19577	No	g15214.t1
MH	Autosomes	GO:0006974	bp	DNA damage response	1	0.30174	0.39882	No	g15214.t1
MH	Autosomes	GO:0034453	bp	microtubule anchoring	1	0.04884	0.14701	No	g15214.t1
MH	Autosomes	GO:0061101	bp	neuroendocrine cell differentiation	1	0.01655	0.10706	No	g15214.t1
MH	Autosomes	GO:0090307	bp	mitotic spindle assembly	1	0.08773	0.18914	No	g15214.t1

MH	Autosomes	GO:0005968	cc	Rab-protein geranylgeranyltransferase complex	1	0.01655	0.10706	No	g15252.t1
MH	Autosomes	GO:0004663	mf	Rab geranylgeranyltransferase activity	1	0.00831	0.09717	No	g15252.t1
MH	Autosomes	GO:0018344	bp	protein geranylgeranylation	1	0.02473	0.11389	No	g15252.t1
MH	Autosomes	GO:0005576	cc	extracellular region	3	1	1	No	g16025.t1,g20983.t1,g21004.t1
MH	Autosomes	GO:0005581	cc	collagen trimer	1	0.30755	0.40474	No	g16025.t1
MH	Autosomes	GO:0005179	mf	hormone activity	1	0.27193	0.37071	No	g16025.t1
MH	Autosomes	GO:0007605	bp	sensory perception of sound	1	0.34692	0.434	No	g16026.t1
MH	Autosomes	GO:0005856	cc	cytoskeleton	1	1	1	No	g16165.t1
MH	Autosomes	GO:0005515	mf	protein binding	2	0.29703	0.39431	No	g16165.t1,g19200.t1
MH	Autosomes	GO:0004725	mf	protein tyrosine phosphatase activity	1	0.45672	0.53607	No	g16166.t1
MH	Autosomes	GO:0017017	mf	MAP kinase tyrosine/serine/threonine phosphatase activity	1	0.10283	0.203	No	g16166.t1
MH	Autosomes	GO:0017018	mf	myosin phosphatase activity	1	0.29588	0.39431	No	g16166.t1
MH	Autosomes	GO:0035335	bp	peptidyl-tyrosine dephosphorylation	1	0.47461	0.55493	No	g16166.t1
MH	Autosomes	GO:0004577	mf	N-acetylglucosaminyldiphosphodolichol N-acetylglucosaminyltransferase activity	2	6.87E-05	0.00696	**Yes**	g17881.t1,g17882.t1
MH	Autosomes	GO:0004843	mf	cysteine-type deubiquitinase activity	1	0.4335	0.51883	No	g17881.t1
MH	Autosomes	GO:0006488	bp	dolichol-linked	2	0.0014	0.08536	No	g17881.t1,g17882.t1

				oligosaccharide biosynthetic process					
MH	Autosomes	GO:0042995	cc	cell projection	2	0.09148	0.19448	No	g17884.t1,g22003.t1
MH	Autosomes	GO:0035556	bp	intracellular signal transduction	2	0.65361	0.71218	No	g17884.t1,g18480.t1
MH	Autosomes	GO:0019934	bp	cGMP-mediated signaling	1	0.05675	0.15403	No	g17950.t1
MH	Autosomes	GO:0032098	bp	regulation of appetite	1	0.02473	0.11389	No	g17950.t1
MH	Autosomes	GO:0007187	bp	G protein-coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	1	0.07237	0.17187	No	g17950.t1
MH	Autosomes	GO:0007268	bp	chemical synaptic transmission	1	0.284	0.38372	No	g17950.t1
MH	Autosomes	GO:0045600	bp	positive regulation of fat cell differentiation	1	0.08008	0.17769	No	g17950.t1
MH	Autosomes	GO:0007631	bp	feeding behavior	1	0.08008	0.17769	No	g17950.t1
MH	Autosomes	GO:0043397	bp	regulation of corticotropin-releasing hormone secretion	1	0.00831	0.09717	No	g17950.t1
MH	Autosomes	GO:0045202	cc	synapse	2	0.34273	0.43054	No	g17950.t1,g22003.t1
MH	Autosomes	GO:0001662	bp	behavioral fear response	1	0.10283	0.203	No	g17950.t1
MH	Autosomes	GO:0051378	mf	serotonin binding	1	0.03283	0.12172	No	g17950.t1
MH	Autosomes	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	2	1	1	No	g17950.t1,g21005.t1
MH	Autosomes	GO:0030425	cc	dendrite	2	0.1574	0.25185	No	g17950.t1,g21001.t1
MH	Autosomes	GO:0001587	mf	Gq/11-coupled serotonin receptor activity	1	0.01655	0.10706	No	g17950.t1
MH	Autosomes	GO:0007626	bp	locomotory behavior	1	0.23449	0.33776	No	g17950.t1
MH	Autosomes	GO:0007208	bp	phospholipase C-activating serotonin receptor signaling	1	0.01655	0.10706	No	g17950.t1

				pathway					
MH	Autosomes	GO:0051482	bp	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G protein-coupled signaling pathway	1	0.07237	0.17187	No	g17950.t1
MH	Autosomes	GO:0070374	bp	positive regulation of ERK1 and ERK2 cascade	1	0.29588	0.39431	No	g17950.t1
MH	Autosomes	GO:0098666	cc	G protein-coupled serotonin receptor complex	1	0.02473	0.11389	No	g17950.t1
MH	Autosomes	GO:0030594	mf	neurotransmitter receptor activity	1	0.06459	0.1664	No	g17950.t1
MH	Autosomes	GO:0071886	mf	1-(4-iodo-2,5-dimethoxyphenyl)propan-2-amine binding	1	0.00831	0.09717	No	g17950.t1
MH	Autosomes	GO:0010513	bp	positive regulation of phosphatidylinositol biosynthetic process	1	0.01655	0.10706	No	g17950.t1
MH	Autosomes	GO:0051209	bp	release of sequestered calcium ion into cytosol	1	0.11769	0.21076	No	g17950.t1
MH	Autosomes	GO:0031644	bp	regulation of nervous system process	1	0.00831	0.09717	No	g17950.t1
MH	Autosomes	GO:0062023	cc	collagen-containing extracellular matrix	1	0.3631	0.44689	No	g17966.t1
MH	Autosomes	GO:0005509	mf	calcium ion binding	2	0.44764	0.52951	No	g17966.t1,g22003.t1
MH	Autosomes	GO:0017095	mf	heparan sulfate 6-O-sulfotransferase activity	1	0.01655	0.10706	No	g17966.t1
MH	Autosomes	GO:0001569	bp	branching involved in blood vessel morphogenesis	2	0.00429	0.09717	No	g17966.t1,g21006.t1

MH	Autosomes	GO:0015015	bp	heparan sulfate proteoglycan biosynthetic process, enzymatic modification	1	0.03283	0.12172	No	g17966.t1
MH	Autosomes	GO:0048048	bp	embryonic eye morphogenesis	1	0.03283	0.12172	No	g17966.t1
MH	Autosomes	GO:0060173	bp	limb development	1	0.11769	0.21076	No	g17966.t1
MH	Autosomes	GO:0030426	cc	growth cone	1	0.22159	0.32387	No	g18481.t1
MH	Autosomes	GO:0016810	mf	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	1	0.08008	0.17769	No	g18481.t1
MH	Autosomes	GO:0031005	mf	filamin binding	1	0.04884	0.14701	No	g18481.t1
MH	Autosomes	GO:0010975	bp	regulation of neuron projection development	1	0.11769	0.21076	No	g18481.t1
MH	Autosomes	GO:0030334	bp	regulation of cell migration	1	0.17472	0.2669	No	g18481.t1
MH	Autosomes	GO:0048678	bp	response to axon injury	1	0.05675	0.15403	No	g18481.t1
MH	Autosomes	GO:0051017	bp	actin filament bundle assembly	1	0.14668	0.23845	No	g18481.t1
MH	Autosomes	GO:0051491	bp	positive regulation of filopodium assembly	1	0.09531	0.19577	No	g18481.t1
MH	Autosomes	GO:0000209	bp	protein polyubiquitination	1	0.27193	0.37071	No	g19198.t1
MH	Autosomes	GO:0006511	bp	ubiquitin-dependent protein catabolic process	1	0.53274	0.60884	No	g19198.t1
MH	Autosomes	GO:0007283	bp	spermatogenesis	1	0.37367	0.45438	No	g19198.t1
MH	Autosomes	GO:0030154	bp	cell differentiation	2	0.23555	0.33776	No	g19198.t1,g19200.t1
MH	Autosomes	GO:0000978	mf	RNA polymerase II cis-regulatory region sequence-specific DNA binding	1	0.72745	0.7898	No	g19199.t1
MH	Autosomes	GO:0000981	mf	DNA-binding transcription factor activity, RNA polymerase II-specific	1	1	1	No	g19199.t1

MH	Autosomes	GO:0035370	cc	UBC13-UEV1A complex	1	0.00831	0.09717	No	g19200.t1
MH	Autosomes	GO:0000151	cc	ubiquitin ligase complex	1	0.17472	0.2669	No	g19200.t1
MH	Autosomes	GO:0045893	bp	positive regulation of DNA-templated transcription	1	0.61788	0.67567	No	g19200.t1
MH	Autosomes	GO:0006282	bp	regulation of DNA repair	1	0.04087	0.13218	No	g19200.t1
MH	Autosomes	GO:0043123	bp	positive regulation of I-kappaB kinase/NF-kappaB signaling	1	0.31332	0.40512	No	g19200.t1
MH	Autosomes	GO:1902523	bp	positive regulation of protein K63-linked ubiquitination	1	0.01655	0.10706	No	g19200.t1
MH	Autosomes	GO:0051092	bp	positive regulation of NF-kappaB transcription factor activity	1	0.23449	0.33776	No	g19200.t1
MH	Autosomes	GO:0006301	bp	postreplication repair	1	0.03283	0.12172	No	g19200.t1
MH	Autosomes	GO:0000381	bp	regulation of alternative mRNA splicing, via spliceosome	2	0.00764	0.09717	No	g19201.t1,g21006.t1
MH	Autosomes	GO:0016607	cc	nuclear speck	3	0.08951	0.19162	No	g19201.t1,g20993.t1,g21006.t1
MH	Autosomes	GO:0006376	bp	mRNA splice site selection	1	0.06459	0.1664	No	g19201.t1
MH	Autosomes	GO:0003729	mf	mRNA binding	2	0.12889	0.21979	No	g19201.t1,g21006.t1
MH	Autosomes	GO:0006355	bp	regulation of DNA-templated transcription	1	0.52562	0.60756	No	g19205.t1
MH	Autosomes	GO:0000421	cc	autophagosome membrane	1	0.10283	0.203	No	g19206.t1
MH	Autosomes	GO:0044754	cc	autolysosome	1	0.04087	0.13218	No	g19206.t1
MH	Autosomes	GO:0005543	mf	phospholipid binding	1	0.24718	0.34719	No	g19206.t1
MH	Autosomes	GO:0008017	mf	microtubule binding	1	1	1	No	g19206.t1
MH	Autosomes	GO:0031625	mf	ubiquitin protein ligase binding	1	0.53274	0.60884	No	g19206.t1
MH	Autosomes	GO:0000045	bp	autophagosome assembly	1	0.12503	0.21474	No	g19206.t1

MH	Autosomes	GO:0000422	bp	autophagy of mitochondrion	1	0.07237	0.17187	No	g19206.t1
MH	Autosomes	GO:0006995	bp	cellular response to nitrogen starvation	1	0.02473	0.11389	No	g19206.t1
MH	Autosomes	GO:0097352	bp	autophagosome maturation	1	0.10283	0.203	No	g19206.t1
MH	Autosomes	GO:0005868	cc	cytoplasmic dynein complex	1	0.08008	0.17769	No	g19207.t1
MH	Autosomes	GO:0005874	cc	microtubule	1	1	1	No	g19207.t1
MH	Autosomes	GO:0045505	mf	dynein intermediate chain binding	1	0.16779	0.26025	No	g19207.t1
MH	Autosomes	GO:0007018	bp	microtubule-based movement	1	0.40433	0.48584	No	g19207.t1
MH	Autosomes	GO:0017150	mf	tRNA dihydrouridine synthase activity	1	0.03283	0.12172	No	g20981.t1
MH	Autosomes	GO:0050660	mf	flavin adenine dinucleotide binding	1	0.31332	0.40512	No	g20981.t1
MH	Autosomes	GO:0002943	bp	tRNA dihydrouridine synthesis	1	0.03283	0.12172	No	g20981.t1
MH	Autosomes	GO:0008083	mf	growth factor activity	1	0.52881	0.60884	No	g20983.t1
MH	Autosomes	GO:0030116	mf	glial cell-derived neurotrophic factor receptor binding	1	0.01655	0.10706	No	g20983.t1
MH	Autosomes	GO:0030971	mf	receptor tyrosine kinase binding	1	0.17472	0.2669	No	g20983.t1
MH	Autosomes	GO:0048731	bp	system development	1	0.09531	0.19577	No	g20983.t1
MH	Autosomes	GO:0006406	bp	mRNA export from nucleus	1	0.12503	0.21474	No	g20993.t1
MH	Autosomes	GO:0043086	bp	negative regulation of catalytic activity	1	0.25345	0.35343	No	g20993.t1
MH	Autosomes	GO:1990830	bp	cellular response to leukemia inhibitory factor	1	0.22807	0.33174	No	g20993.t1
MH	Autosomes	GO:0048024	bp	regulation of mRNA	1	0.04884	0.14701	No	g20993.t1

				splicing, via spliceosome					
MH	Autosomes	GO:0043274	mf	phospholipase binding	1	0.03283	0.12172	No	g20993.t1
MH	Autosomes	GO:0004860	mf	protein kinase inhibitor activity	1	0.08008	0.17769	No	g20993.t1
MH	Autosomes	GO:0006468	bp	protein phosphorylation	1	0.53753	0.61202	No	g20994.t1
MH	Autosomes	GO:0016042	bp	lipid catabolic process	1	0.35236	0.439	No	g20999.t1
MH	Autosomes	GO:0051260	bp	protein homooligomerization	3	0.01558	0.10706	No	g21000.t1,g21001.t1,g21002.t1
MH	Autosomes	GO:0032809	cc	neuronal cell body membrane	1	0.05675	0.15403	No	g21001.t1
MH	Autosomes	GO:0043194	cc	axon initial segment	1	0.03283	0.12172	No	g21001.t1
MH	Autosomes	GO:0043204	cc	perikaryon	1	0.26582	0.36732	No	g21001.t1
MH	Autosomes	GO:0043679	cc	axon terminus	1	0.04087	0.13218	No	g21001.t1
MH	Autosomes	GO:0044224	cc	juxtaparanode region of axon	1	0.01655	0.10706	No	g21001.t1
MH	Autosomes	GO:0005251	mf	delayed rectifier potassium channel activity	1	0.08008	0.17769	No	g21001.t1
MH	Autosomes	GO:0014059	bp	regulation of dopamine secretion	1	0.05675	0.15403	No	g21001.t1
MH	Autosomes	GO:0021633	bp	optic nerve structural organization	1	0.00831	0.09717	No	g21001.t1
MH	Autosomes	GO:0045188	bp	regulation of circadian sleep/wake cycle, non-REM sleep	1	0.00831	0.09717	No	g21001.t1
MH	Autosomes	GO:0007586	bp	digestion	1	0.07237	0.17187	No	g21003.t1
MH	Autosomes	GO:0008028	mf	monocarboxylic acid transmembrane transporter activity	1	0.02473	0.11389	No	g21005.t1
MH	Autosomes	GO:0015718	bp	monocarboxylic acid transport	1	0.04087	0.13218	No	g21005.t1
MH	Autosomes	GO:0055085	bp	transmembrane transport	1	1	1	No	g21005.t1
MH	Autosomes	GO:0036396	cc	RNA N6-methyladenosine methyltransferase	1	0.03283	0.12172	No	g21006.t1

				complex					
MH	Autosomes	GO:0000122	bp	negative regulation of transcription by RNA polymerase II	1	1	1	No	g21006.t1
MH	Autosomes	GO:0001510	bp	RNA methylation	1	0.12503	0.21474	No	g21006.t1
MH	Autosomes	GO:0007221	bp	positive regulation of transcription of Notch receptor target	1	0.02473	0.11389	No	g21006.t1
MH	Autosomes	GO:0038163	bp	thrombopoietin-mediated signaling pathway	1	0.00831	0.09717	No	g21006.t1
MH	Autosomes	GO:0045638	bp	negative regulation of myeloid cell differentiation	1	0.04087	0.13218	No	g21006.t1
MH	Autosomes	GO:0045652	bp	regulation of megakaryocyte differentiation	1	0.00831	0.09717	No	g21006.t1
MH	Autosomes	GO:0048536	bp	spleen development	1	0.10283	0.203	No	g21006.t1
MH	Autosomes	GO:0060412	bp	ventricular septum morphogenesis	1	0.08773	0.18914	No	g21006.t1
MH	Autosomes	GO:0000139	cc	Golgi membrane	1	1	1	No	g22003.t1
MH	Autosomes	GO:0005783	cc	endoplasmic reticulum	1	1	1	No	g22003.t1
MH	Autosomes	GO:0005886	cc	plasma membrane	3	0.02748	0.12107	No	g22003.t1,g22004.t1,g22358.t1
MH	Autosomes	GO:0001540	mf	amyloid-beta binding	1	0.18158	0.276	No	g22003.t1
MH	Autosomes	GO:0019894	mf	kinesin binding	1	0.1323	0.21979	No	g22003.t1
MH	Autosomes	GO:0042988	mf	X11-like protein binding	1	0.02473	0.11389	No	g22003.t1
MH	Autosomes	GO:0007156	bp	homophilic cell adhesion via plasma membrane adhesion molecules	1	0.57388	0.64066	No	g22003.t1
MH	Autosomes	GO:0014065	bp	phosphatidylinositol 3-kinase signaling	1	0.1323	0.21979	No	g22004.t1
MH	Autosomes	GO:1905278	bp	positive regulation of epithelial tube formation	1	0.01655	0.10706	No	g22004.t1

MH	Autosomes	GO:0038089	bp	positive regulation of cell migration by vascular endothelial growth factor signaling pathway	1	0.00831	0.09717	No	g22004.t1
MH	Autosomes	GO:0033031	bp	positive regulation of neutrophil apoptotic process	1	0.02473	0.11389	No	g22004.t1
MH	Autosomes	GO:0036092	bp	phosphatidylinositol-3-phosphate biosynthetic process	1	0.11029	0.20956	No	g22004.t1
MH	Autosomes	GO:0046934	mf	phosphatidylinositol-4,5-bisphosphate 3-kinase activity	1	0.03283	0.12172	No	g22004.t1
MH	Autosomes	GO:0005943	cc	phosphatidylinositol 3-kinase complex, class IA	1	0.04087	0.13218	No	g22004.t1
MH	Autosomes	GO:0010628	bp	positive regulation of gene expression	1	0.58446	0.6461	No	g22004.t1
MH	Autosomes	GO:0001938	bp	positive regulation of endothelial cell proliferation	1	0.11029	0.20956	No	g22004.t1
MH	Autosomes	GO:0052812	mf	phosphatidylinositol-3,4-bisphosphate 5-kinase activity	1	0.03283	0.12172	No	g22004.t1
MH	Autosomes	GO:0010595	bp	positive regulation of endothelial cell migration	1	0.11769	0.21076	No	g22004.t1
MH	Autosomes	GO:0035005	mf	1-phosphatidylinositol-4-phosphate 3-kinase activity	1	0.07237	0.17187	No	g22004.t1
MH	Autosomes	GO:0051897	bp	positive regulation of protein kinase B signaling	1	0.16081	0.2533	No	g22004.t1
MH	Autosomes	GO:0016303	mf	1-phosphatidylinositol-3-kinase activity	1	0.07237	0.17187	No	g22004.t1
MH	Autosomes	GO:0005764	cc	lysosome	1	0.4429	0.52595	No	g22357.t1

MH	Autosomes	GO:0004419	mf	hydroxymethylglutaryl-CoA lyase activity	1	0.01655	0.10706	No	g22357.t1
MH	Autosomes	GO:0004560	mf	alpha-L-fucosidase activity	1	0.01655	0.10706	No	g22357.t1
MH	Autosomes	GO:0006004	bp	fucose metabolic process	1	0.03283	0.12172	No	g22357.t1
MH	Autosomes	GO:0006629	bp	lipid metabolic process	1	0.39933	0.48173	No	g22357.t1
MH	Autosomes	GO:0016139	bp	glycoside catabolic process	1	0.02473	0.11389	No	g22357.t1
MH	Autosomes	GO:0004949	mf	cannabinoid receptor activity	1	0.02473	0.11389	No	g22358.t1
MH	Autosomes	GO:0038171	bp	cannabinoid signaling pathway	1	0.02473	0.11389	No	g22358.t1
MH	Autosomes	GO:0005504	mf	fatty acid binding	1	0.05675	0.15403	No	g22359.t1
MH	Autosomes	GO:0015908	bp	fatty acid transport	1	0.07237	0.17187	No	g22359.t1
MH	Z	GO:0016020	cc	membrane	1	0.42283	0.45711	No	g10113.t1
MH	Z	GO:0004888	mf	transmembrane signaling receptor activity	1	0.05802	0.07735	No	g10113.t1
MH	Z	GO:0007165	bp	signal transduction	2	0.04256	0.06306	No	g10113.t1,g11353.t1
MH	Z	GO:0003964	mf	RNA-directed DNA polymerase activity	10	1.68E-05	0.00023	**Yes**	g10114.t1,g10116.t1,g10117.t1,g11124.t1,g11154.t1,g11158.t1,g11164.t1,g11167.t1,g11352.t1,g11354.t1
MH	Z	GO:0006278	bp	RNA-templated DNA biosynthetic process	10	1.69E-05	0.00023	**Yes**	g10114.t1,g10116.t1,g10117.t1,g11124.t1,g11154.t1,g11158.t1,g11164.t1,g11167.t1,g11352.t1,g11354.t1
MH	Z	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	2	0.76165	0.80173	No	g11157.t1,g11158.t1
MH	Z	GO:0031227	cc	obsolete intrinsic component of endoplasmic reticulum membrane	1	0.00089	0.00509	**Yes**	g11157.t1
MH	Z	GO:0031624	mf	ubiquitin conjugating enzyme binding	1	0.01152	0.02303	**Yes**	g11157.t1

MH	Z	GO:0046872	mf	metal ion binding	1	1	1	No	g11157.t1
MH	Z	GO:0061630	mf	ubiquitin protein ligase activity	1	0.09759	0.11829	No	g11157.t1
MH	Z	GO:0000209	bp	protein polyubiquitination	1	0.03331	0.05329	No	g11157.t1
MH	Z	GO:0030534	bp	adult behavior	1	0.00887	0.01971	**Yes**	g11157.t1
MH	Z	GO:0031398	bp	positive regulation of protein ubiquitination	1	0.01679	0.0292	**Yes**	g11157.t1
MH	Z	GO:0032436	bp	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	1	0.01854	0.0309	**Yes**	g11157.t1
MH	Z	GO:0042415	bp	norepinephrine metabolic process	1	0.00089	0.00509	**Yes**	g11157.t1
MH	Z	GO:0042428	bp	serotonin metabolic process	2	2.26E-06	9.06E-05	**Yes**	g11157.t1,g11159.t1
MH	Z	GO:0050790	bp	regulation of catalytic activity	1	0.22331	0.25521	No	g11157.t1
MH	Z	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	1	0.24677	0.27419	No	g11159.t1
MH	Z	GO:0030425	cc	dendrite	1	0.07305	0.09425	No	g11159.t1
MH	Z	GO:0045202	cc	synapse	1	0.12231	0.14389	No	g11159.t1
MH	Z	GO:0004993	mf	G protein-coupled serotonin receptor activity	1	0.01416	0.02574	**Yes**	g11159.t1
MH	Z	GO:0030594	mf	neurotransmitter receptor activity	1	0.0071	0.01894	**Yes**	g11159.t1
MH	Z	GO:0051378	mf	serotonin binding	1	0.00356	0.01016	**Yes**	g11159.t1
MH	Z	GO:0090722	mf	receptor-receptor interaction	1	0.00178	0.00712	**Yes**	g11159.t1
MH	Z	GO:0001662	bp	behavioral fear response	1	0.01152	0.02303	**Yes**	g11159.t1
MH	Z	GO:0007187	bp	G protein-coupled receptor signaling pathway, coupled to cyclic	1	0.00799	0.01971	**Yes**	g11159.t1

				nucleotide second messenger					
MH	Z	GO:0007198	bp	adenylate cyclase-inhibiting serotonin receptor signaling pathway	1	0.00178	0.00712	**Yes**	g11159.t1
MH	Z	GO:0007214	bp	gamma-aminobutyric acid signaling pathway	1	0.01416	0.02574	**Yes**	g11159.t1
MH	Z	GO:0007268	bp	chemical synaptic transmission	1	0.03503	0.0539	No	g11159.t1
MH	Z	GO:0014062	bp	regulation of serotonin secretion	1	0.00089	0.00509	**Yes**	g11159.t1
MH	Z	GO:0019229	bp	regulation of vasoconstriction	1	0.00887	0.01971	**Yes**	g11159.t1
MH	Z	GO:0035640	bp	exploration behavior	1	0.00267	0.00971	**Yes**	g11159.t1
MH	Z	GO:0042053	bp	regulation of dopamine metabolic process	1	0.00178	0.00712	**Yes**	g11159.t1
MH	Z	GO:0046883	bp	regulation of hormone secretion	1	0.00356	0.01016	**Yes**	g11159.t1
MH	Z	GO:0050795	bp	regulation of behavior	1	0.00356	0.01016	**Yes**	g11159.t1
MH	Z	GO:0005737	cc	cytoplasm	1	1	1	No	g11353.t1
MH	Z	GO:0098978	cc	glutamatergic synapse	1	0.07553	0.09441	No	g11353.t1
MH	Z	GO:0008081	mf	phosphoric diester hydrolase activity	2	4.96E-05	0.0005	**Yes**	g11353.t1,g11355.t1
MH	Z	GO:0016042	bp	lipid catabolic process	1	0.04531	0.06473	No	g11353.t1
MH	Z	GO:0006629	bp	lipid metabolic process	1	0.05295	0.07304	No	g11355.t1

Table S1.9: GO Term enrichment analyses for 173 Autosome genes and 13 Z-linked genes under PH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be FDR- adjusted p values.

Categories:	DNA Information Processing	Cellular structure, metabolism and function	Behaviour	Individual's well-being and health (e.g.sleep,appetite, mood, energy)	Neurotransmission	Fertilization & embryonic development	Immune response

Windows type	Chromosome	GO ID	Ontology	GO Name	Number of Genes	p-value	Adjusted p-value	Significance	Genes
PH	Autosomes	GO:0003964	mf	RNA-directed DNA polymerase activity	32	0.001	0.1009	No	g306.t1,g737.t1,g748.t1,g759.t1,g763.t1,g859.t1,g860.t1,g911.t1,g914.t1,g915.t1,g925.t1,g974.t1,g4475.t1,g5760.t1,g5988.t1,g5990.t1,g5991.t1,g6000.t1,g6006.t1,g6007.t1,g6086.t1,g6700.t1,g8933.t1,g8958.t1,g9010.t1,g9029.t1,g9048.t1,g9050.t1,g12311.t1,g12312.t1,g18136.t1,g18921.t1
PH	Autosomes	GO:0006278	bp	RNA-templated DNA biosynthetic process	32	0.001	0.1009	No	g306.t1,g737.t1,g748.t1,g759.t1,g763.t1,g859.t1,g860.t1,g911.t1,g914.t1,g915.t1,g925.t1,g974.t1,g4475.t1,g5760.t1,g5988.t1,g5990.t1,g5991.t1,g6000.t1,g6006.t1,g6007.t1,g6086.t1,g6700.t1,g8933.t1,g8958.t1,g9010.t1,g9029.t1,g9048.t1,g9050.t1,g12311.t1,g12312.t1,g18136.t1,g18921.t1
PH	Autosomes	GO:0016301	mf	kinase activity	2	0.685	0.7686	No	g733.t1,g4477.t1

PH	Autosomes	GO:0016310	bp	phosphorylation	2	1	1	No	g733.t1,g4477.t1
PH	Autosomes	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	24	0.599	0.6932	No	g736.t1,g756.t1,g860.t1,g6005.t1,g6086.t1,g6517.t1,g7026.t1,g7046.t1,g7057.t1,g7059.t1,g7060.t1,g8931.t1,g9050.t1,g9084.t1,g11855.t1,g13136.t1,g14823.t1,g17187.t1,g18924.t1,g20256.t1,g22281.t1,g22283.t1,g22284.t1,g22286.t1
PH	Autosomes	GO:0000276	cc	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	1	0.111	0.2324	No	g746.t1
PH	Autosomes	GO:0015078	mf	proton transmembrane transporter activity	1	0.118	0.2412	No	g746.t1
PH	Autosomes	GO:0006811	bp	monoatomic ion transport	1	0.25	0.3693	No	g746.t1
PH	Autosomes	GO:0015986	bp	proton motive force-driven ATP synthesis	1	0.131	0.2574	No	g746.t1
PH	Autosomes	GO:0016020	cc	membrane	4	1	1	No	g755.t1,g911.t1,g914.t1,g8323.t1
PH	Autosomes	GO:0005886	cc	plasma membrane	6	0.589	0.6851	No	g756.t1,g6005.t1,g6086.t1,g7046.t1,g8931.t1,g22284.t1
PH	Autosomes	GO:0005509	mf	calcium ion binding	4	0.785	0.8492	No	g756.t1,g6005.t1,g7251.t1,g12444.t1
PH	Autosomes	GO:0007156	bp	homophilic cell adhesion via plasma membrane adhesion molecules	3	0.039	0.1316	No	g756.t1,g6005.t1,g22286.t1
PH	Autosomes	GO:0003824	mf	catalytic activity	4	0.788	0.8492	No	g757.t1,g924.t1,g2141.t1,g13220.t1
PH	Autosomes	GO:0009987	bp	cellular process	3	0.075	0.1818	No	g757.t1,g3922.t1,g8324.t1
PH	Autosomes	GO:0050896	bp	response to stimulus	1	0.293	0.4067	No	g757.t1
PH	Autosomes	GO:0004190	mf	aspartic-type endopeptidase activity	1	0.239	0.3665	No	g765.t1

PH	Autosomes	GO:0006508	bp	proteolysis	3	0.754	0.8262	No	g765.t1,g7039.t1,g22283.t1
PH	Autosomes	GO:0110165	cc	cellular anatomical entity	2	0.244	0.3665	No	g768.t1,g13137.t1
PH	Autosomes	GO:0008076	cc	voltage-gated potassium channel complex	1	0.381	0.4988	No	g861.t1
PH	Autosomes	GO:0005249	mf	voltage-gated potassium channel activity	1	0.412	0.5106	No	g861.t1
PH	Autosomes	GO:0034765	bp	regulation of monoatomic ion transmembrane transport	1	0.581	0.68	No	g861.t1
PH	Autosomes	GO:0071805	bp	potassium ion transmembrane transport	1	0.628	0.7145	No	g861.t1
PH	Autosomes	GO:0008233	mf	peptidase activity	2	0.029	0.1257	No	g911.t1,g22283.t1
PH	Autosomes	GO:0046872	mf	metal ion binding	4	0.199	0.3344	No	g911.t1,g6517.t1,g8956.t1,g12452.t1
PH	Autosomes	GO:0019538	bp	protein metabolic process	1	0.043	0.1316	No	g911.t1
PH	Autosomes	GO:0030234	mf	enzyme regulator activity	1	0.091	0.2032	No	g926.t1
PH	Autosomes	GO:0002080	cc	acrosomal membrane	1	0.029	0.1257	No	g2771.t1
PH	Autosomes	GO:0005615	cc	extracellular space	1	0.377	0.497	No	g2771.t1
PH	Autosomes	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	2	1	1	No	g2771.t1,g14782.t1
PH	Autosomes	GO:0042995	cc	cell projection	1	0.362	0.4806	No	g2771.t1
PH	Autosomes	GO:0043204	cc	perikaryon	1	0.239	0.3665	No	g2771.t1
PH	Autosomes	GO:0005536	mf	glucose binding	1	0.057	0.16	No	g2771.t1
PH	Autosomes	GO:0033300	mf	dehydroascorbic acid transmembrane transporter activity	1	0.015	0.1009	No	g2771.t1
PH	Autosomes	GO:0055056	mf	D-glucose transmembrane transporter activity	1	0.007	0.1009	No	g2771.t1
PH	Autosomes	GO:0046323	bp	glucose import	1	0.007	0.1009	No	g2771.t1
PH	Autosomes	GO:0070837	bp	dehydroascorbic acid transport	1	0.015	0.1009	No	g2771.t1

PH	Autosomes	GO:0005634	cc	nucleus	7	0.271	0.3877	No	g2773.t1,g2774.t1,g7058.t1,g14822.t1,g17032.t1,g17189.t1,g22282.t1
PH	Autosomes	GO:0000981	mf	DNA-binding transcription factor activity, RNA polymerase II-specific	2	0.671	0.7593	No	g2773.t1,g2774.t1
PH	Autosomes	GO:0003677	mf	DNA binding	2	0.333	0.4453	No	g2773.t1,g14822.t1
PH	Autosomes	GO:0006357	bp	regulation of transcription by RNA polymerase II	2	1	1	No	g2773.t1,g2774.t1
PH	Autosomes	GO:0000978	mf	RNA polymerase II cis-regulatory region sequence-specific DNA binding	1	1	1	No	g2774.t1
PH	Autosomes	GO:1902459	bp	positive regulation of stem cell population maintenance	1	0.036	0.1257	No	g2774.t1
PH	Autosomes	GO:0016787	mf	hydrolase activity	1	1	1	No	g3922.t1
PH	Autosomes	GO:0006810	bp	transport	1	0.085	0.1967	No	g3922.t1
PH	Autosomes	GO:0004518	mf	nuclease activity	2	0.023	0.1257	No	g4475.t1,g5988.t1
PH	Autosomes	GO:0005488	mf	binding	2	0.175	0.3066	No	g4475.t1,g5988.t1
PH	Autosomes	GO:0022857	mf	transmembrane transporter activity	2	0.172	0.3046	No	g6086.t1,g11855.t1
PH	Autosomes	GO:0015718	bp	monocarboxylic acid transport	1	0.036	0.1257	No	g6086.t1
PH	Autosomes	GO:0055085	bp	transmembrane transport	3	0.252	0.3704	No	g6086.t1,g11855.t1,g20259.t1
PH	Autosomes	GO:0031624	mf	ubiquitin conjugating enzyme binding	1	0.091	0.2032	No	g6517.t1
PH	Autosomes	GO:0061630	mf	ubiquitin protein ligase activity	1	0.572	0.6766	No	g6517.t1
PH	Autosomes	GO:0031966	cc	mitochondrial membrane	1	0.244	0.3665	No	g6517.t1
PH	Autosomes	GO:0000151	cc	ubiquitin ligase complex	1	0.156	0.2855	No	g6517.t1
PH	Autosomes	GO:0000209	bp	protein polyubiquitination	1	0.244	0.3665	No	g6517.t1

PH	Autosomes	GO:0043066	bp	negative regulation of apoptotic process	2	0.152	0.2852	No	g6517.t1,g7059.t1
PH	Autosomes	GO:0032436	bp	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	1	0.143	0.2772	No	g6517.t1
PH	Autosomes	GO:0016874	mf	ligase activity	2	0.035	0.1257	No	g6517.t1,g12452.t1
PH	Autosomes	GO:0004315	mf	3-oxoacyl-[acyl-carrier-protein] synthase activity	1	0.015	0.1009	No	g6701.t1
PH	Autosomes	GO:0006633	bp	fatty acid biosynthetic process	1	0.233	0.3665	No	g6701.t1
PH	Autosomes	GO:0004222	mf	metalloendopeptidase activity	2	0.242	0.3665	No	g7039.t1,g7059.t1
PH	Autosomes	GO:0008270	mf	zinc ion binding	1	0.196	0.3327	No	g7039.t1
PH	Autosomes	GO:0004930	mf	G protein-coupled receptor activity	2	1	1	No	g7046.t1,g22284.t1
PH	Autosomes	GO:0007186	bp	G protein-coupled receptor signaling pathway	3	0.762	0.8307	No	g7046.t1,g8931.t1,g14782.t1
PH	Autosomes	GO:0000062	mf	fatty-acyl-CoA binding	1	0.036	0.1257	No	g7057.t1
PH	Autosomes	GO:0030242	bp	autophagy of peroxisome	1	0.007	0.1009	No	g7057.t1
PH	Autosomes	GO:0005815	cc	microtubule organizing center	1	0.266	0.3853	No	g7058.t1
PH	Autosomes	GO:0004674	mf	protein serine/threonine kinase activity	1	1	1	No	g7058.t1
PH	Autosomes	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	1	1	1	No	g7058.t1
PH	Autosomes	GO:0005524	mf	ATP binding	8	0.864	0.9269	No	g7058.t1,g7059.t1,g14284.t1, g17186.t1,g18919.t1,g18923. t1,g20256.t1,g20259.t1
PH	Autosomes	GO:0106310	mf	protein serine kinase activity	1	0.614	0.7063	No	g7058.t1
PH	Autosomes	GO:0006468	bp	protein phosphorylation	1	0.525	0.6321	No	g7058.t1

PH	Autosomes	GO:0005743	cc	mitochondrial inner membrane	1	0.408	0.5106	No	g7059.t1
PH	Autosomes	GO:0016604	cc	nuclear body	1	0.394	0.5069	No	g7059.t1
PH	Autosomes	GO:0004176	mf	ATP-dependent peptidase activity	1	0.036	0.1257	No	g7059.t1
PH	Autosomes	GO:0016887	mf	ATP hydrolysis activity	3	0.438	0.5361	No	g7059.t1,g18919.t1,g20256.t1
PH	Autosomes	GO:0006515	bp	protein quality control for misfolded or incompletely synthesized proteins	1	0.036	0.1257	No	g7059.t1
PH	Autosomes	GO:0008283	bp	cell population proliferation	1	0.244	0.3665	No	g7059.t1
PH	Autosomes	GO:0034214	bp	protein hexamerization	1	0.036	0.1257	No	g7059.t1
PH	Autosomes	GO:0034982	bp	mitochondrial protein processing	1	0.015	0.1009	No	g7059.t1
PH	Autosomes	GO:0035694	bp	mitochondrial protein catabolic process	1	0.022	0.1256	No	g7059.t1
PH	Autosomes	GO:0051015	mf	actin filament binding	1	0.581	0.68	No	g7251.t1
PH	Autosomes	GO:0005546	mf	phosphatidylinositol-4,5-bisphosphate binding	1	0.131	0.2574	No	g7251.t1
PH	Autosomes	GO:0051016	bp	barbed-end actin filament capping	1	0.078	0.183	No	g7251.t1
PH	Autosomes	GO:0051014	bp	actin filament severing	1	0.043	0.1316	No	g7251.t1
PH	Autosomes	GO:0005902	cc	microvillus	1	0.124	0.252	No	g7251.t1
PH	Autosomes	GO:0005737	cc	cytoplasm	3	0.01	0.1009	No	g7251.t1,g14284.t1,g17186.t1
PH	Autosomes	GO:0045010	bp	actin nucleation	1	0.078	0.183	No	g7251.t1
PH	Autosomes	GO:0015629	cc	actin cytoskeleton	1	0.293	0.4067	No	g7251.t1
PH	Autosomes	GO:0001726	cc	ruffle	1	0.216	0.3546	No	g7251.t1
PH	Autosomes	GO:0032433	cc	filopodium tip	1	0.036	0.1257	No	g7251.t1
PH	Autosomes	GO:0030027	cc	lamellipodium	2	0.074	0.1818	No	g7251.t1,g12452.t1
PH	Autosomes	GO:0043229	cc	intracellular organelle	1	0.105	0.2232	No	g8324.t1

PH	Autosomes	GO:0045177	cc	apical part of cell	1	0.162	0.2891	No	g8931.t1
PH	Autosomes	GO:0048471	cc	perinuclear region of cytoplasm	2	0.323	0.4375	No	g8931.t1,g12452.t1
PH	Autosomes	GO:0016499	mf	orexin receptor activity	1	0.015	0.1009	No	g8931.t1
PH	Autosomes	GO:0007631	bp	feeding behavior	1	0.071	0.177	No	g8931.t1
PH	Autosomes	GO:0022410	bp	circadian sleep/wake cycle process	1	0.015	0.1009	No	g8931.t1
PH	Autosomes	GO:0030971	mf	receptor tyrosine kinase binding	1	0.156	0.2855	No	g8956.t1
PH	Autosomes	GO:0001525	bp	angiogenesis	1	0.313	0.4324	No	g8956.t1
PH	Autosomes	GO:0030154	bp	cell differentiation	1	0.559	0.6652	No	g8956.t1
PH	Autosomes	GO:0003676	mf	nucleic acid binding	2	1	1	No	g9030.t1,g16295.t1
PH	Autosomes	GO:0016746	mf	acyltransferase activity	1	0.143	0.2772	No	g9084.t1
PH	Autosomes	GO:0005730	cc	nucleolus	2	0.702	0.7808	No	g11855.t1,g14821.t1
PH	Autosomes	GO:0032040	cc	small-subunit processome	1	0.15	0.2841	No	g11855.t1
PH	Autosomes	GO:0042254	bp	ribosome biogenesis	2	0.008	0.1009	No	g11855.t1,g14821.t1
PH	Autosomes	GO:0004843	mf	cysteine-type deubiquitinase activity	1	0.394	0.5069	No	g12246.t1
PH	Autosomes	GO:0016579	bp	protein deubiquitination	1	0.272	0.3877	No	g12246.t1
PH	Autosomes	GO:0022625	cc	cytosolic large ribosomal subunit	1	0.18	0.3137	No	g12248.t1
PH	Autosomes	GO:0003735	mf	structural constituent of ribosome	1	0.522	0.6317	No	g12248.t1
PH	Autosomes	GO:0000463	bp	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA)	1	0.071	0.177	No	g12248.t1
PH	Autosomes	GO:0005544	mf	calcium-dependent phospholipid binding	1	0.192	0.3291	No	g12444.t1
PH	Autosomes	GO:0016491	mf	oxidoreductase activity	1	0.515	0.6267	No	g12451.t1
PH	Autosomes	GO:0005794	cc	Golgi apparatus	2	0.687	0.7686	No	g12452.t1,g16322.t1
PH	Autosomes	GO:0009966	bp	regulation of signal	1	0.239	0.3665	No	g12452.t1

				transduction					
PH	Autosomes	GO:0016567	bp	protein ubiquitination	1	1	1	No	g12452.t1
PH	Autosomes	GO:0000139	cc	Golgi membrane	1	1	1	No	g13136.t1
PH	Autosomes	GO:0007030	bp	Golgi organization	2	0.027	0.1257	No	g13136.t1,g22282.t1
PH	Autosomes	GO:0004197	mf	cysteine-type endopeptidase activity	1	0.162	0.2891	No	g13137.t1
PH	Autosomes	GO:0051603	bp	proteolysis involved in protein catabolic process	1	0.064	0.1724	No	g13137.t1
PH	Autosomes	GO:0016459	cc	myosin complex	1	0.328	0.4413	No	g14284.t1
PH	Autosomes	GO:0003774	mf	cytoskeletal motor activity	1	0.277	0.3926	No	g14284.t1
PH	Autosomes	GO:0003779	mf	actin binding	1	1	1	No	g14284.t1
PH	Autosomes	GO:0044298	cc	cell body membrane	1	0.015	0.1009	No	g14782.t1
PH	Autosomes	GO:1990913	cc	sperm head plasma membrane	1	0.015	0.1009	No	g14782.t1
PH	Autosomes	GO:0005502	mf	11-cis retinal binding	1	0.029	0.1257	No	g14782.t1
PH	Autosomes	GO:0008020	mf	G protein-coupled photoreceptor activity	1	0.043	0.1316	No	g14782.t1
PH	Autosomes	GO:0007601	bp	visual perception	1	0.403	0.5089	No	g14782.t1
PH	Autosomes	GO:0007603	bp	phototransduction, visible light	1	0.022	0.1256	No	g14782.t1
PH	Autosomes	GO:0007634	bp	optokinetic behavior	1	0.007	0.1009	No	g14782.t1
PH	Autosomes	GO:0016038	bp	absorption of visible light	1	0.029	0.1257	No	g14782.t1
PH	Autosomes	GO:0042752	bp	regulation of circadian rhythm	1	0.131	0.2574	No	g14782.t1
PH	Autosomes	GO:0043052	bp	thermotaxis	1	0.015	0.1009	No	g14782.t1
PH	Autosomes	GO:0050960	bp	detection of temperature stimulus involved in thermoception	1	0.015	0.1009	No	g14782.t1
PH	Autosomes	GO:0071482	bp	cellular response to light stimulus	1	0.043	0.1316	No	g14782.t1
PH	Autosomes	GO:1990384	bp	hyaloid vascular plexus regression	1	0.015	0.1009	No	g14782.t1

PH	Autosomes	GO:0005525	mf	GTP binding	1	0.727	0.8041	No	g14821.t1
PH	Autosomes	GO:0005930	cc	axoneme	1	0.156	0.2855	No	g16294.t1
PH	Autosomes	GO:0003341	bp	cilium movement	1	0.036	0.1257	No	g16294.t1
PH	Autosomes	GO:0036159	bp	inner dynein arm assembly	1	0.036	0.1257	No	g16294.t1
PH	Autosomes	GO:0005768	cc	endosome	1	0.399	0.5089	No	g16322.t1
PH	Autosomes	GO:0016324	cc	apical plasma membrane	1	0.403	0.5089	No	g16322.t1
PH	Autosomes	GO:0030139	cc	endocytic vesicle	1	0.105	0.2232	No	g16322.t1
PH	Autosomes	GO:0098981	cc	cholinergic synapse	1	0.022	0.1256	No	g16322.t1
PH	Autosomes	GO:0004630	mf	phospholipase D activity	1	0.007	0.1009	No	g16322.t1
PH	Autosomes	GO:0035091	mf	phosphatidylinositol binding	1	0.323	0.4375	No	g16322.t1
PH	Autosomes	GO:0070290	mf	N-acylphosphatidylethanolamine-specific phospholipase D activity	1	0.007	0.1009	No	g16322.t1
PH	Autosomes	GO:0006654	bp	phosphatidic acid biosynthetic process	1	0.05	0.1465	No	g16322.t1
PH	Autosomes	GO:0016042	bp	lipid catabolic process	1	0.318	0.4364	No	g16322.t1
PH	Autosomes	GO:0032534	bp	regulation of microvillus assembly	1	0.015	0.1009	No	g16322.t1
PH	Autosomes	GO:0045727	bp	positive regulation of translation	2	0.008	0.1009	No	g16322.t1,g17189.t1
PH	Autosomes	GO:0048017	bp	inositol lipid-mediated signaling	1	0.022	0.1256	No	g16322.t1
PH	Autosomes	GO:0048870	bp	cell motility	1	0.071	0.177	No	g16322.t1
PH	Autosomes	GO:0098693	bp	regulation of synaptic vesicle cycle	1	0.029	0.1257	No	g16322.t1
PH	Autosomes	GO:0004829	mf	threonine-tRNA ligase activity	1	0.029	0.1257	No	g17186.t1
PH	Autosomes	GO:0006435	bp	threonyl-tRNA aminoacylation	1	0.029	0.1257	No	g17186.t1
PH	Autosomes	GO:0046331	bp	lateral inhibition	1	0.007	0.1009	No	g17187.t1
PH	Autosomes	GO:0004000	mf	adenosine deaminase	1	0.043	0.1316	No	g17188.t1

				activity					
PH	Autosomes	GO:0006154	bp	adenosine catabolic process	1	0.015	0.1009	No	g17188.t1
PH	Autosomes	GO:0046103	bp	inosine biosynthetic process	1	0.007	0.1009	No	g17188.t1
PH	Autosomes	GO:0005844	cc	polysome	1	0.078	0.183	No	g17189.t1
PH	Autosomes	GO:0017022	mf	myosin binding	1	0.057	0.16	No	g17189.t1
PH	Autosomes	GO:0035613	mf	RNA stem-loop binding	1	0.036	0.1257	No	g17189.t1
PH	Autosomes	GO:0048027	mf	mRNA 5'-UTR binding	1	0.064	0.1724	No	g17189.t1
PH	Autosomes	GO:1990825	mf	sequence-specific mRNA binding	1	0.022	0.1256	No	g17189.t1
PH	Autosomes	GO:0006396	bp	RNA processing	1	0.227	0.3647	No	g17189.t1
PH	Autosomes	GO:0032967	bp	positive regulation of collagen biosynthetic process	1	0.05	0.1465	No	g17189.t1
PH	Autosomes	GO:1902416	bp	positive regulation of mRNA binding	1	0.022	0.1256	No	g17189.t1
PH	Autosomes	GO:0051536	mf	iron-sulfur cluster binding	1	0.057	0.16	No	g18919.t1
PH	Autosomes	GO:0016226	bp	iron-sulfur cluster assembly	1	0.091	0.2032	No	g18919.t1
PH	Autosomes	GO:0030173	cc	obsolete integral component of Golgi membrane	1	0.15	0.2841	No	g18920.t1
PH	Autosomes	GO:0009306	bp	protein secretion	1	0.111	0.2324	No	g18920.t1
PH	Autosomes	GO:0016192	bp	vesicle-mediated transport	1	0.437	0.5361	No	g18920.t1
PH	Autosomes	GO:0006355	bp	regulation of DNA-templated transcription	1	0.73	0.8041	No	g18923.t1
PH	Autosomes	GO:0006303	bp	double-strand break repair via nonhomologous end joining	1	0.071	0.177	No	g20258.t1
PH	Autosomes	GO:0034707	cc	chloride channel complex	1	0.227	0.3647	No	g20259.t1
PH	Autosomes	GO:0005254	mf	chloride channel activity	1	0.186	0.3215	No	g20259.t1
PH	Autosomes	GO:0140359	mf	ABC-type transporter activity	1	0.227	0.3647	No	g20259.t1

PH	Autosomes	GO:0006749	bp	glutathione metabolic process	1	0.162	0.2891	No	g20259.t1
PH	Autosomes	GO:0006821	bp	chloride transport	1	0.266	0.3853	No	g20259.t1
PH	Autosomes	GO:0010467	bp	gene expression	1	0.105	0.2232	No	g20259.t1
PH	Autosomes	GO:0042101	cc	T cell receptor complex	1	0.029	0.1257	No	g22281.t1
PH	Autosomes	GO:0004888	mf	transmembrane signaling receptor activity	1	0.39	0.5069	No	g22281.t1
PH	Autosomes	GO:0007166	bp	cell surface receptor signaling pathway	1	0.412	0.5106	No	g22281.t1
PH	Autosomes	GO:0000922	cc	spindle pole	1	0.216	0.3546	No	g22282.t1
PH	Autosomes	GO:0005811	cc	lipid droplet	1	0.204	0.3408	No	g22282.t1
PH	Autosomes	GO:0005828	cc	kinetochore microtubule	1	0.064	0.1724	No	g22282.t1
PH	Autosomes	GO:0005829	cc	cytosol	2	0.036	0.1257	No	g22282.t1,g22286.t1
PH	Autosomes	GO:0070939	cc	Dsl1/NZR complex	1	0.015	0.1009	No	g22282.t1
PH	Autosomes	GO:1990423	cc	RZZ complex	1	0.015	0.1009	No	g22282.t1
PH	Autosomes	GO:0000132	bp	establishment of mitotic spindle orientation	1	0.05	0.1465	No	g22282.t1
PH	Autosomes	GO:0006888	bp	endoplasmic reticulum to Golgi vesicle-mediated transport	1	0.25	0.3693	No	g22282.t1
PH	Autosomes	GO:0007080	bp	mitotic metaphase plate congression	1	0.091	0.2032	No	g22282.t1
PH	Autosomes	GO:0007094	bp	mitotic spindle assembly checkpoint signaling	1	0.071	0.177	No	g22282.t1
PH	Autosomes	GO:0007096	bp	regulation of exit from mitosis	1	0.015	0.1009	No	g22282.t1
PH	Autosomes	GO:0034501	bp	protein localization to kinetochore	1	0.043	0.1316	No	g22282.t1
PH	Autosomes	GO:0051301	bp	cell division	1	0.543	0.6492	No	g22282.t1
PH	Autosomes	GO:0065003	bp	protein-containing complex assembly	1	0.118	0.2412	No	g22282.t1
PH	Autosomes	GO:0045202	cc	synapse	1	1	1	No	g22284.t1

PH	Autosomes	GO:0004952	mf	dopamine neurotransmitter receptor activity	1	0.036	0.1257	No	g22284.t1
PH	Autosomes	GO:0001963	bp	synaptic transmission, dopaminergic	1	0.071	0.177	No	g22284.t1
PH	Autosomes	GO:0007195	bp	adenylate cyclase-inhibiting dopamine receptor signaling pathway	1	0.029	0.1257	No	g22284.t1
PH	Autosomes	GO:0065008	bp	regulation of biological quality	1	0.098	0.2157	No	g22284.t1
PH	Autosomes	GO:1903530	bp	regulation of secretion by cell	1	0.007	0.1009	No	g22284.t1
PH	Autosomes	GO:0009897	cc	external side of plasma membrane	1	0.293	0.4067	No	g22286.t1
PH	Autosomes	GO:0005515	mf	protein binding	1	0.62	0.7089	No	g22286.t1
PH	Autosomes	GO:0099560	bp	synaptic membrane adhesion	1	0.043	0.1316	No	g22286.t1
PH	Z	GO:0005634	cc	nucleus	1	0.57	0.5829	No	g11037.t1
PH	Z	GO:0005829	cc	cytosol	1	0.429	0.4486	No	g11037.t1
PH	Z	GO:0005854	cc	nascent polypeptide-associated complex	1	6E-04	0.0036	**Yes**	g11037.t1
PH	Z	GO:0042788	cc	polysomal ribosome	1	0.007	0.0135	**Yes**	g11037.t1
PH	Z	GO:0005515	mf	protein binding	1	0.07	0.0918	No	g11037.t1
PH	Z	GO:0001701	bp	in utero embryonic development	1	0.007	0.0135	**Yes**	g11037.t1
PH	Z	GO:0015031	bp	protein transport	1	0.073	0.0939	No	g11037.t1
PH	Z	GO:1905551	bp	negative regulation of protein localization to endoplasmic reticulum	1	6E-04	0.0036	**Yes**	g11037.t1
PH	Z	GO:0003779	mf	actin binding	1	0.077	0.0952	No	g11047.t1
PH	Z	GO:0007399	bp	nervous system development	1	0.036	0.0491	**Yes**	g11047.t1
PH	Z	GO:0001669	cc	acrosomal vesicle	1	0.007	0.0137	**Yes**	g11048.t1

PH	Z	GO:0005615	cc	extracellular space	1	0.198	0.2217	No	g11048.t1
PH	Z	GO:0016020	cc	membrane	1	0.288	0.3085	No	g11048.t1
PH	Z	GO:0042582	cc	azurophil granule	1	6E-04	0.0036	**Yes**	g11048.t1
PH	Z	GO:0060473	cc	cortical granule	1	0.002	0.0063	**Yes**	g11048.t1
PH	Z	GO:1905379	cc	beta-N-acetylhexosaminidase complex	1	6E-04	0.0036	**Yes**	g11048.t1
PH	Z	GO:0004563	mf	beta-N-acetylhexosaminidase activity	1	0.002	0.0072	**Yes**	g11048.t1
PH	Z	GO:0008375	mf	acetylglucosaminyltransferase activity	1	0.003	0.0084	**Yes**	g11048.t1
PH	Z	GO:0042802	mf	identical protein binding	1	0.167	0.1957	No	g11048.t1
PH	Z	GO:0102148	mf	N-acetyl-beta-D-galactosaminidase activity	1	0.002	0.0072	**Yes**	g11048.t1
PH	Z	GO:0001501	bp	skeletal system development	1	0.012	0.0191	**Yes**	g11048.t1
PH	Z	GO:0006689	bp	ganglioside catabolic process	1	0.002	0.0063	**Yes**	g11048.t1
PH	Z	GO:0006874	bp	intracellular calcium ion homeostasis	1	0.008	0.0146	**Yes**	g11048.t1
PH	Z	GO:0007040	bp	lysosome organization	1	0.008	0.0146	**Yes**	g11048.t1
PH	Z	GO:0007341	bp	penetration of zona pellucida	1	6E-04	0.0036	**Yes**	g11048.t1
PH	Z	GO:0007605	bp	sensory perception of sound	1	0.028	0.0387	**Yes**	g11048.t1
PH	Z	GO:0007626	bp	locomotory behavior	1	0.017	0.0268	**Yes**	g11048.t1
PH	Z	GO:0008049	bp	male courtship behavior	1	6E-04	0.0036	**Yes**	g11048.t1
PH	Z	GO:0008360	bp	regulation of cell shape	1	0.022	0.0314	**Yes**	g11048.t1
PH	Z	GO:0008654	bp	phospholipid biosynthetic process	1	0.005	0.0135	**Yes**	g11048.t1
PH	Z	GO:0009313	bp	oligosaccharide catabolic process	1	0.001	0.0056	**Yes**	g11048.t1

PH	Z	GO:0019915	bp	lipid storage	1	0.006	0.0135	**Yes**	g11048.t1
PH	Z	GO:0030203	bp	glycosaminoglycan metabolic process	1	0.002	0.0063	**Yes**	g11048.t1
PH	Z	GO:0042552	bp	myelination	1	0.007	0.0135	**Yes**	g11048.t1
PH	Z	GO:0043615	bp	astrocyte cell migration	1	6E-04	0.0036	**Yes**	g11048.t1
PH	Z	GO:0045944	bp	positive regulation of transcription by RNA polymerase II	1	0.17	0.1957	No	g11048.t1
PH	Z	GO:0048477	bp	oogenesis	1	0.004	0.0104	**Yes**	g11048.t1
PH	Z	GO:0050885	bp	neuromuscular process controlling balance	1	0.009	0.015	**Yes**	g11048.t1
PH	Z	GO:0016301	mf	kinase activity	2	0.007	0.0135	**Yes**	g11410.t1,g11411.t1
PH	Z	GO:0016773	mf	phosphotransferase activity, alcohol group as acceptor	1	0.004	0.0104	**Yes**	g11410.t1
PH	Z	GO:0046496	bp	nicotinamide nucleotide metabolic process	1	0.001	0.0056	**Yes**	g11410.t1
PH	Z	GO:0016310	bp	phosphorylation	1	0.141	0.1706	No	g11411.t1
PH	Z	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	2	1	1	No	g11412.t1,g11415.t1
PH	Z	GO:0005509	mf	calcium ion binding	1	0.239	0.2621	No	g11414.t1
PH	Z	GO:0004896	mf	cytokine receptor activity	1	0.009	0.0153	**Yes**	g11415.t1
PH	Z	GO:0019221	bp	cytokine-mediated signaling pathway	1	0.019	0.0276	**Yes**	g11415.t1

Table S1.10: KEGG enrichment analyses for 112 Autosome genes and 14 Z-linked genes under LH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be Benjamini-Hochberg adjusted p values.

Windows Type	Chomosome Type	KEGG Pathway	P-value	Adjusted P-value	Significant	Genes
LH	Autosomes	Glycosaminoglycan biosynthesis	**0.031**	0.758	No	NDST2, HS3ST1
LH	Autosomes	Riboflavin metabolism	**0.040**	0.758	No	ACP1
LH	Autosomes	Thiamine metabolism	0.075	0.758	No	ACP1
LH	Autosomes	Glycerophospholipid metabolism	0.091	0.758	No	DGKA, TAZ
LH	Autosomes	Steroid biosynthesis	0.098	0.758	No	LIPA
LH	Autosomes	Folate biosynthesis	0.126	0.758	No	PTS
LH	Autosomes	Biosynthesis of unsaturated fatty acids	0.130	0.758	No	SCD
LH	Autosomes	Lysosome	0.141	0.758	No	LIPA, LGMN
LH	Autosomes	Citrate cycle (TCA cycle)	0.144	0.758	No	SDHD
LH	Autosomes	Propanoate metabolism	0.161	0.758	No	ACSS3
LH	Autosomes	Tyrosine metabolism	0.170	0.758	No	TPO
LH	Autosomes	DNA replication	0.170	0.758	No	RNASEH1
LH	Autosomes	African trypanosomiasis	0.174	0.758	No	IL18
LH	Autosomes	Thyroid cancer	0.174	0.758	No	NCOA4
LH	Autosomes	Ferroptosis	0.191	0.758	No	NCOA4
LH	Autosomes	Ribosome	0.196	0.758	No	RPL18A, MRPL32
LH	Autosomes	Porphyrin and chlorophyll metabolism	0.199	0.758	No	BLVRA
LH	Autosomes	ABC transporters	0.208	0.758	No	ABCG2
LH	Autosomes	Malaria	0.228	0.758	No	IL18
LH	Autosomes	Cholesterol metabolism	0.228	0.758	No	LIPA
LH	Autosomes	Autoimmune thyroid disease	0.240	0.758	No	TPO
LH	Autosomes	Legionellosis	0.255	0.758	No	IL18
LH	Autosomes	Glycerolipid metabolism	0.271	0.758	No	DGKA
LH	Autosomes	Lysine degradation	0.278	0.758	No	ASH1L

LH	Autosomes	Cytosolic DNA-sensing pathway	0.278	0.758	No	IL18
LH	Autosomes	Inflammatory bowel disease	0.285	0.758	No	IL18
LH	Autosomes	Adherens junction	0.307	0.758	No	ACP1
LH	Autosomes	PPAR signaling pathway	0.318	0.758	No	SCD
LH	Autosomes	Thyroid hormone synthesis	0.322	0.758	No	TPO
LH	Autosomes	Antigen processing and presentation	0.332	0.758	No	LGMN
LH	Autosomes	Taste transduction	0.359	0.758	No	PKD2L1
LH	Autosomes	Cardiac muscle contraction	0.362	0.758	No	TRDN
LH	Autosomes	ECM-receptor interaction	0.366	0.758	No	GP1BB
LH	Autosomes	Bile secretion	0.372	0.758	No	ABCG2
LH	Autosomes	Rheumatoid arthritis	0.382	0.758	No	IL18
LH	Autosomes	TGF-beta signaling pathway	0.385	0.758	No	BAMBI
LH	Autosomes	Phosphatidylinositol signaling system	0.395	0.758	No	DGKA
LH	Autosomes	Choline metabolism in cancer	0.398	0.758	No	DGKA
LH	Autosomes	Hematopoietic cell lineage	0.401	0.758	No	GP1BB
LH	Autosomes	Viral protein interaction with cytokine and cytokine receptor	0.404	0.758	No	IL18
LH	Autosomes	Ribosome biogenesis in eukaryotes	0.428	0.778	No	BMS1
LH	Autosomes	Leukocyte transendothelial migration	0.446	0.778	No	CD99
LH	Autosomes	AMPK signaling pathway	0.463	0.778	No	SCD
LH	Autosomes	Platelet activation	0.474	0.778	No	GP1BB
LH	Autosomes	Oocyte meiosis	0.487	0.778	No	SLK
LH	Autosomes	Purine metabolism	0.487	0.778	No	NUDT5
LH	Autosomes	Oxidative phosphorylation	0.498	0.778	No	SDHD
LH	Autosomes	Autophagy	0.508	0.778	No	SH3GLB1
LH	Autosomes	Yersinia infection	0.508	0.778	No	IL18
LH	Autosomes	Cell adhesion molecules	0.536	0.788	No	CD99
LH	Autosomes	Phospholipase D signaling pathway	0.536	0.788	No	DGKA
LH	Autosomes	Non-alcoholic fatty liver disease	0.552	0.796	No	SDHD
LH	Autosomes	Hippo signaling pathway	0.570	0.800	No	WWTR1

LH	Autosomes	Wnt signaling pathway	0.577	0.800	No	BAMBI
LH	Autosomes	Influenza A	0.590	0.800	No	IL18
LH	Autosomes	Tuberculosis	0.607	0.800	No	IL18
LH	Autosomes	NOD-like receptor signaling pathway	0.609	0.800	No	IL18
LH	Autosomes	RNA transport	0.619	0.800	No	XPOT
LH	Autosomes	Pathogenic Escherichia coli infection	0.640	0.805	No	IL18
LH	Autosomes	Diabetic cardiomyopathy	0.651	0.805	No	SDHD
LH	Autosomes	Lipid and atherosclerosis	0.672	0.805	No	IL18
LH	Autosomes	Thermogenesis	0.700	0.805	No	SDHD
LH	Autosomes	Coronavirus disease	0.700	0.805	No	RPL18A
LH	Autosomes	Calcium signaling pathway	0.713	0.805	No	TRDN
LH	Autosomes	Shigellosis	0.721	0.805	No	IL18
LH	Autosomes	Salmonella infection	0.726	0.805	No	IL18
LH	Autosomes	Parkinson disease	0.726	0.805	No	SDHD
LH	Autosomes	Endocytosis	0.730	0.805	No	SH3GLB1
LH	Autosomes	Prion disease	0.758	0.824	No	SDHD
LH	Autosomes	Cytokine-cytokine receptor interaction	0.784	0.840	No	IL18
LH	Autosomes	Huntington disease	0.797	0.841	No	SDHD
LH	Autosomes	Amyotrophic lateral sclerosis	0.850	0.877	No	SDHD
LH	Autosomes	Alzheimer disease	0.854	0.877	No	SDHD
LH	Autosomes	Pathways of neurodegeneration	0.916	0.929	No	SDHD
LH	Autosomes	Pathways in cancer	0.938	0.938	No	NCOA4
LH	Z	Terpenoid backbone biosynthesis	**0.011**	0.083	No	HMGCR
LH	Z	Tyrosine metabolism	**0.018**	0.083	No	TPO
LH	Z	Autoimmune thyroid disease	**0.026**	0.083	No	TPO
LH	Z	Thyroid hormone synthesis	**0.037**	0.083	No	TPO
LH	Z	Bile secretion	**0.044**	0.083	No	HMGCR
LH	Z	Ribosome biogenesis in eukaryotes	0.053	0.083	No	RPP40
LH	Z	AMPK signaling pathway	0.058	0.083	No	HMGCR
LH	Z	RNA transport	0.089	0.112	No	RPP40

LH	Z	cAMP signaling pathway	0.103	0.114	No	SOX9
LH	Z	Salmonella infection	0.118	0.118	No	PLEKHM1

Table S1.11: KEGG enrichment analyses for 80 Autosome genes and 9 Z-linked genes under MH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be Benjamini-Hochberg adjusted p values.

Windows Type	Chomosome Type	KEGG Pathway	P-value	Adjusted P-value	Significant	Genes
MH	Autosomes	Bladder cancer	**0.011**	0.391	No	RB1,CDKN1A
MH	Autosomes	Melanoma	**0.032**	0.391	No	RB1,CDKN1A
MH	Autosomes	Non-small cell lung cancer	**0.032**	0.391	No	RB1,CDKN1A
MH	Autosomes	Glioma	**0.035**	0.391	No	RB1,CDKN1A
MH	Autosomes	Chronic myeloid leukemia	**0.036**	0.391	No	RB1,CDKN1A
MH	Autosomes	Pancreatic cancer	**0.036**	0.391	No	RB1,CDKN1A
MH	Autosomes	Small cell lung cancer	0.050	0.404	No	RB1,CDKN1A
MH	Autosomes	Human T-cell leukemia virus 1 infection	0.054	0.404	No	RB1,CDKN1A,MMP7
MH	Autosomes	Prostate cancer	0.055	0.404	No	RB1,CDKN1A
MH	Autosomes	Cell cycle	0.084	0.442	No	RB1,CDKN1A
MH	Autosomes	Breast cancer	0.112	0.442	No	RB1,CDKN1A
MH	Autosomes	Gastric cancer	0.115	0.442	No	RB1,CDKN1A
MH	Autosomes	Cushing syndrome	0.123	0.442	No	RB1,CDKN1A
MH	Autosomes	Cellular senescence	0.124	0.442	No	RB1,CDKN1A
MH	Autosomes	Hepatitis C	0.125	0.442	No	RB1,CDKN1A
MH	Autosomes	JAK-STAT signaling pathway	0.132	0.442	No	CDKN1A,CSF2RA
MH	Autosomes	Hepatitis B	0.132	0.442	No	RB1,CDKN1A
MH	Autosomes	Thyroid cancer	0.135	0.442	No	CDKN1A
MH	Autosomes	Hepatocellular carcinoma	0.140	0.442	No	RB1,CDKN1A
MH	Autosomes	Various types of N-glycan biosynthesis	0.141	0.442	No	ALG13
MH	Autosomes	Nicotine addiction	0.145	0.442	No	GRIA4
MH	Autosomes	Neuroactive ligand-receptor interaction	0.148	0.442	No	P2RY8,CNR2, GRIA4
MH	Autosomes	Type II diabetes mellitus	0.165	0.442	No	ADIPOQ
MH	Autosomes	Kaposi sarcoma-associated herpesvirus infection	0.174	0.442	No	RB1,CDKN1A

MH	Autosomes	N-Glycan biosynthesis	0.178	0.442	No	ALG13
MH	Autosomes	Epstein-Barr virus infection	0.186	0.442	No	RB1,CDKN1A
MH	Autosomes	Glycosaminoglycan biosynthesis	0.187	0.442	No	HS6ST2
MH	Autosomes	Viral carcinogenesis	0.188	0.442	No	RB1,CDKN1A
MH	Autosomes	Endometrial cancer	0.203	0.449	No	CDKN1A
MH	Autosomes	Basal cell carcinoma	0.219	0.449	No	CDKN1A
MH	Autosomes	Human cytomegalovirus infection	0.219	0.449	No	RB1,CDKN1A
MH	Autosomes	Adipocytokine signaling pathway	0.237	0.449	No	ADIPOQ
MH	Autosomes	Amphetamine addiction	0.237	0.449	No	GRIA4
MH	Autosomes	Renal cell carcinoma	0.237	0.449	No	CDKN1A
MH	Autosomes	Adherens junction	0.243	0.449	No	SNAI1
MH	Autosomes	p53 signaling pathway	0.249	0.449	No	CDKN1A
MH	Autosomes	PPAR signaling pathway	0.252	0.449	No	ADIPOQ
MH	Autosomes	ErbB signaling pathway	0.283	0.472	No	CDKN1A
MH	Autosomes	Taste transduction	0.286	0.472	No	SCN2A
MH	Autosomes	Colorectal cancer	0.286	0.472	No	CDKN1A
MH	Autosomes	Circadian entrainment	0.316	0.499	No	GRIA4
MH	Autosomes	Hematopoietic cell lineage	0.321	0.499	No	CSF2RA
MH	Autosomes	Longevity regulating pathway	0.329	0.499	No	ADIPOQ
MH	Autosomes	Parathyroid hormone synthesis, secretion and action	0.340	0.499	No	CDKN1A
MH	Autosomes	Pathways in cancer	0.343	0.499	No	RB1,CDKN1A, CSF2RA
MH	Autosomes	HIF-1 signaling pathway	0.348	0.499	No	CDKN1A
MH	Autosomes	Glutamatergic synapse	0.360	0.505	No	GRIA4
MH	Autosomes	Human papillomavirus infection	0.371	0.505	No	RB1,CDKN1A
MH	Autosomes	AMPK signaling pathway	0.375	0.505	No	ADIPOQ
MH	Autosomes	FoxO signaling pathway	0.402	0.523	No	CDKN1A
MH	Autosomes	Dopaminergic synapse	0.404	0.523	No	GRIA4
MH	Autosomes	Retrograde endocannabinoid signaling	0.440	0.547	No	GRIA4
MH	Autosomes	Oxytocin signaling pathway	0.453	0.547	No	CDKN1A

MH	Autosomes	Non-alcoholic fatty liver disease	0.456	0.547	No	ADIPOQ
MH	Autosomes	Ribosome	0.462	0.547	No	MRPL19
MH	Autosomes	Necroptosis	0.464	0.547	No	SPATA2
MH	Autosomes	Wnt signaling pathway	0.479	0.554	No	MMP7
MH	Autosomes	Transcriptional misregulation in cancer	0.529	0.602	No	CDKN1A
MH	Autosomes	Proteoglycans in cancer	0.553	0.619	No	CDKN1A
MH	Autosomes	cAMP signaling pathway	0.572	0.629	No	GRIA4
MH	Autosomes	Chemical carcinogenesis	0.609	0.659	No	RB1
MH	Autosomes	Cytokine-cytokine receptor interaction	0.687	0.727	No	CSF2RA
MH	Autosomes	Huntington disease	0.700	0.727	No	GRIA4
MH	Autosomes	MicroRNAs in cancer	0.705	0.727	No	CDKN1A
MH	Autosomes	PI3K-Akt signaling pathway	0.752	0.764	No	CDKN1A
MH	Autosomes	Pathways of neurodegeneration	0.847	0.847	No	GRIA4
MH	Z	Cardiac muscle contraction	**0.034**	0.098	No	TRDN
MH	Z	Ribosome biogenesis in eukaryotes	**0.042**	0.098	No	RPP40
MH	Z	Autophagy	0.054	0.098	No	DEPTOR
MH	Z	mTOR signaling pathway	0.060	0.098	No	DEPTOR
MH	Z	Hippo signaling pathway	0.063	0.098	No	SCRIB
MH	Z	Tight junction	0.066	0.098	No	SCRIB
MH	Z	RNA transport	0.072	0.098	No	RPP40
MH	Z	Viral carcinogenesis	0.078	0.098	No	SCRIB
MH	Z	Calcium signaling pathway	0.092	0.102	No	TRDN
MH	Z	Human papillomavirus infection	0.125	0.125	No	SCRIB

Table S1.12: KEGG enrichment analyses for 62 Autosome genes and 17 Z-linked genes under PH selection regions. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be Benjamini-Hochberg adjusted p values.

Windows Type	Chomosome Type	KEGG Pathway	P-value	Adjusted P-value	Significant	Genes
PH	Autosomes	Fatty acid biosynthesis	**0.001**	0.128	No	OXSM,CBR4
PH	Autosomes	Primary immunodeficiency	**0.006**	0.281	No	CIITA,CD3D
PH	Autosomes	Cocaine addiction	**0.010**	0.307	No	DRD2,NFKB1
PH	Autosomes	Antigen processing and presentation	**0.024**	0.331	No	CIITA,LGMN
PH	Autosomes	PD-L1 expression and PD-1 checkpoint pathway in cancer	**0.031**	0.331	No	CD3D,NFKB1
PH	Autosomes	Th1 and Th2 cell differentiation	**0.033**	0.331	No	CD3D,NFKB1
PH	Autosomes	IL-17 signaling pathway	**0.034**	0.331	No	MAPK15,NFKB1
PH	Autosomes	Chagas disease	**0.040**	0.331	No	CD3D,NFKB1
PH	Autosomes	T cell receptor signaling pathway	**0.041**	0.331	No	CD3D,NFKB1
PH	Autosomes	Th17 cell differentiation	**0.044**	0.331	No	CD3D,NFKB1
PH	Autosomes	Ribosome biogenesis in eukaryotes	**0.044**	0.331	No	BMS1,RPP40
PH	Autosomes	HIF-1 signaling pathway	**0.045**	0.331	No	ANGPT2,NFKB1
PH	Autosomes	Toxoplasmosis	**0.047**	0.331	No	CIITA,NFKB1
PH	Autosomes	Lysosome	0.060	0.390	No	ABCA2,LGMN
PH	Autosomes	Measles	0.069	0.420	No	CD3D,NFKB1
PH	Autosomes	Neuroactive ligand-receptor interaction	0.089	0.475	No	GRID1,DRD2,HCRTR2
PH	Autosomes	Influenza A	0.100	0.475	No	CIITA,NFKB1
PH	Autosomes	Tuberculosis	0.107	0.475	No	CIITA,NFKB1
PH	Autosomes	Kaposi sarcoma-associated herpesvirus infection	0.120	0.475	No	ANGPT2,NFKB1
PH	Autosomes	Pathogenic Escherichia coli infection	0.125	0.475	No	MYO10,NFKB1
PH	Autosomes	Epstein-Barr virus infection	0.130	0.475	No	CD3D,NFKB1
PH	Autosomes	ABC transporters	0.131	0.475	No	ABCA2
PH	Autosomes	Rap1 signaling pathway	0.138	0.475	No	ANGPT2,DRD2

PH	Autosomes	Human immunodeficiency virus 1 infection	0.140	0.475	No	CD3D,NFKB1
PH	Autosomes	cAMP signaling pathway	0.145	0.475	No	DRD2,NFKB1
PH	Autosomes	Human T-cell leukemia virus 1 infection	0.148	0.475	No	CD3D,NFKB1
PH	Autosomes	Ras signaling pathway	0.162	0.475	No	ANGPT2,NFKB1
PH	Autosomes	Legionellosis	0.162	0.475	No	NFKB1
PH	Autosomes	Cytosolic DNA-sensing pathway	0.178	0.475	No	NFKB1
PH	Autosomes	Inflammatory bowel disease	0.183	0.475	No	NFKB1
PH	Autosomes	Acute myeloid leukemia	0.188	0.475	No	NFKB1
PH	Autosomes	Adipocytokine signaling pathway	0.193	0.475	No	NFKB1
PH	Autosomes	Epithelial cell signaling in Helicobacter pylori infection	0.196	0.475	No	NFKB1
PH	Autosomes	Prolactin signaling pathway	0.196	0.475	No	NFKB1
PH	Autosomes	RIG-I-like receptor signaling pathway	0.196	0.475	No	NFKB1
PH	Autosomes	Chronic myeloid leukemia	0.211	0.475	No	NFKB1
PH	Autosomes	Pancreatic cancer	0.211	0.475	No	NFKB1
PH	Autosomes	Pertussis	0.211	0.475	No	NFKB1
PH	Autosomes	Leishmaniasis	0.213	0.475	No	NFKB1
PH	Autosomes	B cell receptor signaling pathway	0.223	0.475	No	NFKB1
PH	Autosomes	MAPK signaling pathway	0.231	0.475	No	ANGPT2,NFKB1
PH	Autosomes	Taste transduction	0.235	0.475	No	ENTPD2
PH	Autosomes	Gap junction	0.240	0.475	No	DRD2
PH	Autosomes	Small cell lung cancer	0.249	0.475	No	NFKB1
PH	Autosomes	Fc gamma R-mediated phagocytosis	0.261	0.475	No	MYO10
PH	Autosomes	Prostate cancer	0.261	0.475	No	NFKB1
PH	Autosomes	Hematopoietic cell lineage	0.265	0.475	No	CD3D
PH	Autosomes	AGE-RAGE signaling pathway in diabetic complications	0.267	0.475	No	NFKB1
PH	Autosomes	Amoebiasis	0.272	0.475	No	NFKB1
PH	Autosomes	Longevity regulating pathway	0.272	0.475	No	NFKB1
PH	Autosomes	Toll-like receptor signaling pathway	0.277	0.475	No	NFKB1
PH	Autosomes	C-type lectin receptor signaling pathway	0.277	0.475	No	NFKB1
PH	Autosomes	NF-kappa B signaling pathway	0.277	0.475	No	NFKB1

PH	Autosomes	Insulin resistance	0.286	0.481	No	NFKB1
PH	Autosomes	TNF signaling pathway	0.294	0.486	No	NFKB1
PH	Autosomes	PI3K-Akt signaling pathway	0.300	0.486	No	ANGPT2,NFKB1
PH	Autosomes	Sphingolipid signaling pathway	0.310	0.486	No	NFKB1
PH	Autosomes	Neurotrophin signaling pathway	0.310	0.486	No	NFKB1
PH	Autosomes	Osteoclast differentiation	0.327	0.494	No	NFKB1
PH	Autosomes	Purine metabolism	0.331	0.494	No	ENTPD2
PH	Autosomes	Relaxin signaling pathway	0.331	0.494	No	NFKB1
PH	Autosomes	Dopaminergic synapse	0.337	0.495	No	DRD2
PH	Autosomes	Yersinia infection	0.347	0.500	No	NFKB1
PH	Autosomes	Fluid shear stress and atherosclerosis	0.351	0.500	No	NFKB1
PH	Autosomes	Apoptosis	0.358	0.501	No	NFKB1
PH	Autosomes	Non-alcoholic fatty liver disease	0.383	0.518	No	NFKB1
PH	Autosomes	Cellular senescence	0.385	0.518	No	NFKB1
PH	Autosomes	Hepatitis C	0.387	0.518	No	NFKB1
PH	Autosomes	Hepatitis B	0.397	0.523	No	NFKB1
PH	Autosomes	Protein processing in endoplasmic reticulum	0.413	0.537	No	DNAJC1
PH	Autosomes	NOD-like receptor signaling pathway	0.431	0.540	No	NFKB1
PH	Autosomes	Alcoholism	0.440	0.540	No	DRD2
PH	Autosomes	RNA transport	0.440	0.540	No	RPP40
PH	Autosomes	Neutrophil extracellular trap formation	0.445	0.540	No	NFKB1
PH	Autosomes	Transcriptional misregulation in cancer	0.451	0.540	No	NFKB1
PH	Autosomes	Chemokine signaling pathway	0.451	0.540	No	NFKB1
PH	Autosomes	Viral carcinogenesis	0.469	0.547	No	NFKB1
PH	Autosomes	Diabetic cardiomyopathy	0.469	0.547	No	NFKB1
PH	Autosomes	Lipid and atherosclerosis	0.489	0.563	No	NFKB1
PH	Autosomes	Human cytomegalovirus infection	0.505	0.574	No	NFKB1
PH	Autosomes	Coronavirus disease	0.515	0.579	No	NFKB1
PH	Autosomes	Chemical carcinogenesis	0.526	0.579	No	NFKB1
PH	Autosomes	Shigellosis	0.536	0.579	No	NFKB1

PH	Autosomes	Salmonella infection	0.541	0.579	No	NFKB1
PH	Autosomes	Parkinson disease	0.541	0.579	No	DRD2
PH	Autosomes	MicroRNAs in cancer	0.621	0.657	No	NFKB1
PH	Autosomes	Human papillomavirus infection	0.645	0.675	No	NFKB1
PH	Autosomes	Alzheimer disease	0.685	0.709	No	NFKB1
PH	Autosomes	Pathways of neurodegeneration	0.775	0.793	No	NFKB1
PH	Autosomes	Herpes simplex virus 1 infection	0.791	0.800	No	NFKB1
PH	Autosomes	Pathways in cancer	0.812	0.812	No	NFKB1
PH	Z	**FoxO signaling pathway**	**0.005**	**0.049**	Yes	SKP2,IL7R
PH	Z	**Ubiquitin mediated proteolysis**	**0.006**	**0.049**	Yes	FANCL,SKP2
PH	Z	Nicotinate and nicotinamide metabolism	**0.029**	0.127	No	NADK
PH	Z	Primary immunodeficiency	**0.032**	0.127	No	IL7R
PH	Z	Fanconi anemia pathway	**0.045**	0.144	No	FANCL
PH	Z	Pathways in cancer	0.074	0.162	No	SKP2,IL7R
PH	Z	Small cell lung cancer	0.075	0.162	No	SKP2
PH	Z	Hematopoietic cell lineage	0.081	0.162	No	IL7R
PH	Z	Cell cycle	0.100	0.178	No	SKP2
PH	Z	mTOR signaling pathway	0.123	0.188	No	SKP2
PH	Z	JAK-STAT signaling pathway	0.129	0.188	No	IL7R
PH	Z	Epstein-Barr virus infection	0.159	0.193	No	SKP2
PH	Z	Viral carcinogenesis	0.159	0.193	No	SKP2
PH	Z	cAMP signaling pathway	0.169	0.193	No	SOX9
PH	Z	Cytokine-cytokine receptor interaction	0.223	0.238	No	IL7R
PH	Z	PI3K-Akt signaling pathway	0.262	0.262	No	IL7R

Table S2.1. Total sample size for each sampling location by Hawaiian Island that was genetically vetted by Wells, Lavretsky et al. (2019). The average assignment probability (± Standard Deviation) to either Hawaiian duck or mallard (Feral or Wild) across sampling locations. I also provide the proportion of samples that were identified as native Hawaiian duck (≥ 0.90 assignment probability to Hawaiian duck), Feral mallard (≥ 0.90 assignment probability to Hawaiian Feral mallards), or Wild mallard (≥ 0.90 assignment probability to Wild North American mallards).

Island/Location	Sample Size	Avg. Assignment Probability to Koloa (Standard Deviation)	Avg. Assignment Probability to mallard (Standard Deviation)	Proportion of Pure Hawaiian duck	Proportion of Feral mallard	Proportion of Wild mallard	Proportion of mixed ancestry
(1) Hawaiʻi							
(a) Waimea	9	0.75 (0.081)	0.24 (0.081)	0	0	0	1.00
(b) Hilo	9	0.0023 (0.0057)	0.99 (0.0057)	0	0	0	1.00
(2) Maui	17	0.17 (0.13)	0.82 (0.14)	0	0	0	1.00
(3) Molokaʻi	6	0.18 (0.048)	0.78 (0.05)	0	0	0	1.00
(4) Oʻahu							
(a) Kiʻi	36	0.74 (0.20)	0.26 (0.20)	0	0	0	1.00
(b) Kaneohe	16	0.66 (0.071)	0.34 (0.072)	0	0	0	1.00
(c) Kawainui	24	0.48 (0.17)	0.52 (0.17)	0	0	0	1.00
(d) Hamakua	63	0.28 (0.19)	0.72 (0.19)	0	0.32	0	0.68
(e) Paiko	17	0.012 (0.020)	0.98 (0.022)	0	1.00	0	0
(f) Airport	8	0.38 (0.33)	0.62 (0.33)	0	0.25	0	0.75
(g) Zoo	21	0.31 (0.13)	0.69 (0.13)	0	0	0	1.00
(h) Hawaiikai	25	0.035 (0.068)	0.96 (0.068)	0	0.88	0	0.12
(4) Kauai							
(a) Hanalei	155	0.98 (0.057)	0.017 (0.054)	0.96	0	0	0.04
(b) Huleʻia	18	0.96 (0.030)	0.034 (0.026)	1.00	0	0	0

Table S2.2. Optimized management strategies for three wetlands on O‘ahu of varying starting average Hawaiian duck ancestry (Table 1). Static and adaptive optimizations were done under management strategies including removing only or a combination of supplementation and removal at different levels of Hawaiian duck ancestry in the removed individuals (i.e. 0.1,0.5, 0.9). Success of each strategy was determined by the predicted final ancestry, with the objective set to reach 99% Hawaiian duck ancestry. Note that while static optimization simply allocates the same number of individuals across set generations (gen), adaptive optimization varies the strategy across generations.

Model	**Management Strategy**		**Wetland (Current Avg. Hawaiian duck Ancestry of Population)**		
			Hamukua **-28%**	**Kawainui** **-48%**	**Ki'i** **-74%**
Static Optimization	Removing Only	Ancestry pull = 0.1	Remove 61 Hybrids/gen x 20 gens =**1,220 Hybrids** Final Ancestry = 0.20 = **Failed**	Remove 42 Hybrids /gen x 20 gens = **840 Hybrids** Final Ancestry = 0.55 = **Failed**	Remove 50 Hybrids /gen x 20 gens = **100 Hybrids** Final Ancestry = 0.72 = **Failed**
		Ancestry pull = 0.5	Remove 14 Hybrids/gen x 20 gens = **280 Hybrids** Final Ancestry = 0 = **Failed**	Remove 69 Hybrids /gen x 20 gens = **1,380 Hybrids** Final Ancestry = 0.51 = **Failed**	Remove 56 Hybrids /gen x 20 gens = **100 Hybrids** Final Ancestry = 0.76 = **Failed**
		Ancestry pull = 0.9	Remove 13 Hybrids/gen x 20 gens =**260 Hybrids** Final Ancestry = 0 = **Failed**	Remove 8 Hybrids /gen x 20 gens = **80 Hybrids** Final Ancestry = 0.45 = **Failed**	Remove 15 Hybrids /gen x 20 gens = **100 Hybrids** Final Ancestry = 0.73 = **Failed**
	Supplement + Removing	Ancestry pull = 0.1	Supplement 95 Hawaiian ducks /gen x 20 gens = **1,900 Hawaiian ducks** + Remove 66 Hawaiian ducks /gen x 20 gens = 1,320 Hybrids Final Ancestry = 0.999 = **Achieved**	Supplement 39 Hawaiian ducks /gen x 20 gens = **780 Hawaiian ducks** + Remove 27 Hawaiian ducks /gen x 20 gens = 540 Hybrids Final Ancestry = 0.999 = **Achieved**	Supplement 24 Hawaiian ducks /gen x 20 gens = **480 Hawaiian ducks** + Remove 28 Hawaiian ducks /gen x 20 gens = 560 Hybrids Final Ancestry = 0.999 = **Achieved**

		Ancestry pull = 0.5	Supplement 68 Hawaiian ducks /gen x 20 gens = **1,360 Hawaiian ducks** + Remove 60 Hawaiian ducks /gen x 20 gens = 1,300 Hybrids Final Ancestry = 1.000 = **Achieved**	Supplement 35 Hawaiian ducks /gen x 20 gens = **700 Hawaiian duck**s + Remove 14 Hawaiian ducks /gen x 20 gens = 280 Hybrids Final Ancestry = 0.99 = **Achieved**	Supplement 39 Hawaiian ducks /gen x 20 gens = **780 Hawaiian ducks** + Remove 69 Hawaiian ducks /gen x 20 gens = 1,380 Hybrids Final Ancestry = 0.999 = **Achieved**
		Ancestry pull = 0.9	Supplement 44 Hawaiian ducks /gen x 20 gens = 880 Hawaiian ducks + Remove 37 Hawaiian ducks /gen x 20 gens = 740 Hybrids Final Ancestry = 0.999 = **Achieved**	Supplement 26 Hawaiian ducks /gen x 20 gens = **520 Hawaiian ducks** + Remove 21 Hawaiian ducks /gen x 20 gens = 420 Hybrids Final Ancestry = 0.99 = **Achieved**	Supplement 25 Hawaiian ducks /gen x 20 gens = **500 Hawaiian ducks** + Remove 17 Hawaiian ducks /gen x 20 gens = 340 Hybrids Final Ancestry = 0.999 = **Achieved**
Adaptive Optimization	Removing Only	Ancestry pull = 0.1	Remove 100 total Hybrids Vary strategy for 4 generations Final Ancestry = 0.245 = **Failed**	Remove 100 total Hybrids Vary strategy for 6 generations Final Ancestry = 0.551 = **Failed**	Remove 100 total Hybrids Vary strategy for 7 generations Final Ancestry = 0.756 = **Failed**
		Ancestry pull = 0.5	Remove 100 total Hybrids Vary strategy for 4 generations Final Ancestry = 0.29 = **Failed**	Remove 100 total Hybrids Vary strategy for 5 generations Final Ancestry = 0.483 = **Failed**	Remove 100 total Hybrids Vary strategy for 2 generations Final Ancestry = 0.705 = **Failed**
		Ancestry pull = 0.9	Remove 100 total Hybrids Vary strategy for 2 generations Final Ancestry = 0.311 = **Failed**	Remove 100 total Hybrids Vary strategy for 5 generations Final Ancestry = 0.469 = **Failed**	Remove 100 total Hybrids Vary strategy for 5 generations Final Ancestry = 0.827 = **Failed**
	Supplement + Removing	Ancestry pull = 0.1	Supplement 420 total Hawaiian ducks + Remove 298 total Hybrids Vary strategy for 14 generations Final Ancestry = 1.000 = **Achieved**	Supplement 278 total Hawaiian ducks + Remove 141 total Hybrids Vary strategy for 15 generations Final Ancestry = 1.000 = **Achieved**	Supplement 180 total Hawaiian ducks + Remove 79 total Hybrids Vary strategy for 11 generations Final Ancestry = 1.000 = **Achieved**
		Ancestry pull = 0.5	Supplement 344 total Hawaiian ducks + Remove 195 total Hybrids Vary strategy for 14 generations Final Ancestry = 1.000 = **Achieved**	Supplement 203 total Hawaiian ducks + Remove 119 total Hybrids Vary strategy for 15 generations Final Ancestry = 1.000 = **Achieved**	Supplement 170 total Hawaiian ducks + Remove 70 total Hybrids Vary strategy for 15 generations Final Ancestry = 0.99 = **Achieved**
		Ancestry pull = 0.9	Supplement 284 total Hawaiian ducks + Remove 163 total Hybrids Vary strategy for 11 generations Final Ancestry = 1.000 = **Achieved**	Supplement 175 total Hawaiian ducks + Remove 72 total Hybrids Vary strategy for 14 generations Final Ancestry = 1.000 = **Achieved**	Supplement 111 total Hawaiian ducks + Remove 34 total Hybrids Vary strategy for 8 generations Final Ancestry = 1.000 = **Achieved**

Table S3.1. Sample information.

Species	Museum / Source	Tissue #	State/ Province	Country	Latitude	Longitude	Sex
Anas platyrhynchos	UTEP-BC	PL112316	New Mexico	US	32.9540	-107.2960	M
Anas laysanensis	MWFB	25385-2	Hawaii	US	28.2097	-177.3754	M
Anas wyviliana	MWFB	25386-4	Hawaii	US	22.0362	-159.7789	M
Anas rubripes	UTEP-BC	PL010414	Ohio	US	39.9980	-83.8163	M
Anas diazi	UTEP-BC	PL3014	Chihuahua	Mexico	28.1149	-106.9890	M
Anas superciliosa superciliosa	UTEP-BC	PL8002	Totara Flat	New Zealand	-42.3122	171.6027	M
Anas fulvigula fulvigula	UTEP-BC	TTFL2	Florida	US	26.9630	80.6028	M
Anas fulvigula maculosa	UTEP-BC	TPWD31	Texas	US	28.9183	95.4328	M
Anas undulata	NA	RR15	Free state	South Africa	-27.1383	28.3856	M
Anas sparsa	NA	RR9	Mpumalanga	South Africa	-26.5194	29.7344	M
Anas luzonica	NA	PAN042	Nueva Ecija	Philippines	15.8440	121.1584	M

Table S3.2. Adult survival rate for each duck used to calculate generation time (G). The age of maturity for mallard-like ducks

Species	Average Adult Survival	Generation time	Reference
Wild Mallard	0.57	2.32	Drilling etal. 2020
Laysan Duck	0.906	10.64	Reynolds & Citta 2010
Hawaiian Duck	0.7479	3.96	Malachowski et al. xxx
American Black Duck	0.6305	2.7	Longcore et al. 2020
Mexican duck	0.58	2.72	Used the same information of Mottled duck (GC)
New Zealand Grey Duck	0.56	2.27	Caithness et al. 1991
Mottled duck-(FL)	0.5845	2.4	Johnson et al 1995, Biefield et al. 2020
Mottled duck-(WGC)	0.58	2.7	Johnson (2009); USFW (2014)
Yellow-billed duck	0.698	2.72	Dean & Skead 1989
African Black duck	0.57	2.32	Drilling etal. 2020
Philippine duck	0.7479	2.4	Used the same information of Hawaiian duck

Table S3.3. Genome sequencing results.

	Autosomes				Z- chromosome			
Species	No. reads	Covered bases (Mbp)	Alignment (%)	Mean coverage (x)	No. reads	Covered bases (Mbp)	Alignment (%)	Mean coverage (x)
Laysan duck	23258274	33.96	98.91	83.9	6E+07	78.51	99.13	90.7
Hawaiian duck	22332019	34.02	99.10	84.6	6E+07	78.63	99.28	85.7
Mallard	20777591	34.07	99.31	67.4	5E+07	78.72	99.39	78.1
Mexican duck	5982526	33.91	98.91	27.5	1E+07	78.17	98.69	24.2
American Black duck	5779911	34.03	99.13	24.6	1E+07	78.64	99.29	23.7
(FL) Mottled duck	5074507	34.65	99.11	19.0	1E+07	78.61	99.25	19.4
(WGC) Mottled duck	5085653	34.02	99.27	19.8	1E+07	78.63	99.27	21.3
Philippine duck	7946951	33.98	99.02	32.0	2E+07	78.49	99.09	31.9
New Zealand Grey Duck	4873445	34.00	99.07	20.4	1E+07	78.63	99.28	19.5
African Black duck	4496620	33.94	98.84	17.9	1E+07	78.39	98.98	18.4
Yellow-billed duck	4275047	33.98	99.02	17.0	1E+07	78.51	99.13	16.9

Table S4. List of genes and associated GO Terms in outlier windows per chromosome for top and bottom 1% in autosomes and Z-sex chromosome

Windows Type	Chr	Start	End	Gene	Description	GO Term
Top	chr1	46300001	46400000	g1218.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016020
Top	chr1	46300001	46400000	g1219.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr1	46300001	46400000	g1220.t1	hypothetical protein llap_11333	
Top	chr1	46300001	46400000	g1221.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0016310; GO:0003964; GO:0016301
Top	chr1	47000001	47100000	g1258.t1	hypothetical protein llap_13265	GO:0006508; GO:0008236; GO:0016787
Top	chr1	47000001	47100000	g1259.t1	28S ribosomal protein S12, mitochondrial	GO:0006412; GO:0003735; GO:0015935
Top	chr1	47000001	47100000	g1260.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0003964; GO:0016020; GO:0016021
Top	chr1	47000001	47100000	g1261.t1	hypothetical protein C8258_32110, partial	GO:0006278; GO:0003964
Top	chr1	47000001	47100000	g1262.t1	hypothetical protein llap_9844	GO:0006278; GO:0003964
Top	chr1	47400001	47500000	g1275.t1	hypothetical protein AS27_13285, partial	
Top	chr1	47400001	47500000	g1276.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0003964

Top	chr1	47400001	47500000	g1277.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr1	96200001	96300000	g2358.t1	Protein tyrosine phosphatase-like protein PTPLAD2	GO:0006633; GO:0032508; GO:0003677; GO:0005524; GO:0102158; GO:0102343; GO:0102344; GO:0102345; GO:0005789; GO:0016021
Top	chr1	96200001	96300000	g2359.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr1	120100001	120200000	g2829.t1	potassium voltage-gated channel subfamily A member 1	GO:0034765; GO:0051260; GO:0071805; GO:0005249; GO:0008076
Top	chr1	120100001	120200000	g2830.t1	PREDICTED: DEP domain-containing protein 1B, partial	GO:0006278; GO:0003964
Top	chr1	120100001	120200000	g2831.t1	hypothetical protein	GO:0006278; GO:0003964
Top	chr1	127000001	127100000	g3011.t1	PREDICTED: DEP domain-containing protein 1B, partial	GO:0006278; GO:0003964
Top	chr1	128400001	128500000	g3047.t1	PREDICTED: thyroid peroxidase-like, partial	GO:0003824
Top	chr1	128400001	128500000	g3048.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Top	chr1	128800001	128900000	g3059.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Top	chr1	129300001	129400000	NA		
Top	chr1	129500001	129600000	g3081.t1	cGMP-inhibited 3',5'-cyclic phosphodiesterase A isoform X5	GO:0001556; GO:0009410; GO:0019934; GO:0040020; GO:0043116; GO:0043117; GO:0043951; GO:0060282; GO:0071321; GO:0071560; GO:0004115; GO:0046872; GO:0005829; GO:0016020
Top	chr1	129500001	129600000	g3082.t1	RNA-directed DNA polymerase from mobile	GO:0006278; GO:0003964

					element jockey, partial	
Top	chr1	129500001	129600000	g3083.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr1	147700001	147800000	g3515.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr1	147800001	147900000	g3516.t1	kit ligand isoform 2 precursor	GO:0001934; GO:0002684; GO:0003006; GO:0007155; GO:0008284; GO:0043066; GO:0045597; GO:0048468; GO:0048513; GO:0048731; GO:1902533; GO:0005125; GO:0005173; GO:0005576; GO:0005737; GO:0005886; GO:0016021; GO:0030027; GO:0030175
Top	chr1	147900001	148000000	g3517.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr2	77800001	77900000	g5970.t1	usherin	GO:0110165
Top	chr2	77800001	77900000	g5971.t1	hypothetical protein N324_03341, partial	
Top	chr2	77800001	77900000	g5972.t1	protein N-lysine methyltransferase FAM173B	GO:0018022; GO:0016279; GO:0016021; GO:0031966
Top	chr2	77900001	78000000	NA		
Top	chr2	87000001	87100000	g6140.t1	reverse transcriptase, partial	GO:0006278; GO:0003964
Top	chr2	87000001	87100000	g6141.t1	hypothetical protein DUI87_22201	GO:0006278; GO:0003964
Top	chr2	87000001	87100000	g6142.t1	hypothetical protein Anapl_15300	GO:0046872
Top	chr2	87000001	87100000	g6143.t1	hypothetical protein HGM15179_017134, partial	GO:0006278; GO:0003964; GO:0005488
Top	chr2	87000001	87100000	g6144.t1	reverse transcriptase family protein, partial	GO:0010951; GO:0004867; GO:0005615

Top	chr2	89400001	89500000	g6238.t1	-	
Top	chr2	89400001	89500000	g6239.t1	hypothetical protein llap_14412	GO:0016740
Top	chr2	89400001	89500000	g6240.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Top	chr2	92200001	92300000	g6306.t1	hypothetical protein DV515_00001767	GO:0009117; GO:0016311; GO:0000166; GO:0000287; GO:0106411; GO:0005737
Top	chr2	92200001	92300000	g6307.t1	peptidyl-prolyl cis-trans isomerase FKBP9	GO:0000413; GO:0006457; GO:0003755; GO:0005509; GO:0005783
Top	chr2	92200001	92300000	g6308.t1	hypothetical protein DUI87_18532	GO:0008380
Top	chr2	93700001	93800000	g6335.t1	hypothetical protein DUI87_18684	
Top	chr2	94300001	94400000	g6348.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Top	chr2	94400001	94500000	g6349.t1	hypothetical protein	
Top	chr2	94500001	94600000	NA		
Top	chr2	94600001	94700000	g6350.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Top	chr2	94700001	94800000	g6351.t1	hypothetical protein DUI87_22404	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Top	chr2	94800001	94900000	g6352.t1	inhibin beta A chain	GO:0007165; GO:0005179; GO:0008083; GO:0005576
Top	chr2	94900001	95000000	NA		
Top	chr2	95000001	95100000	NA		
Top	chr2	96800001	96900000	g6380.t1	hypothetical protein llap_4130	
Top	chr2	97500001	97600000	g6394.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr2	97500001	97600000	g6395.t1	hypothetical protein	

					C5E46_35750	
Top	chr2	97500001	97600000	g6396.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chr2	97500001	97600000	g6397.t1	hypothetical protein llap_2540	
Top	chr2	97500001	97600000	g6398.t1	hypothetical protein llap_19921	GO:0006278; GO:0003964
Top	chr2	97700001	97800000	g6402.t1	rna-directed dna polymerase from mobile element hypothetical protein	GO:0006278; GO:0003964
Top	chr2	97700001	97800000	g6403.t1	hypothetical protein HGM15179_019816, partial	GO:0006278; GO:0003964
Top	chr2	97700001	97800000	g6404.t1	death-associated protein 1	GO:0003824
Top	chr2	97700001	97800000	g6405.t1	contactin-associated protein-like 2 isoform X2	GO:0071205; GO:0016021; GO:0033010
Top	chr2	98100001	98200000	g6417.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Top	chr2	98100001	98200000	g6418.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964; GO:0016021
Top	chr4	29000001	29100000	g12198.t1	---NA---	
Top	chr4	35200001	35300000	g12297.t1	reverse transcriptase family protein	GO:0006278; GO:0003964
Top	chr5	20400001	20500000	NA		
Top	chr5	48800001	48900000	g13844.t1	hypothetical protein llap_16894	
Top	chr6	8700001	8800000	NA		
Top	chr6	9200001	9300000	g14795.t1	hypothetical protein AV530_015968	

Top	chr6	9200001	9300000	g14796.t1	choline O-acetyltransferase	GO:0060416; GO:0004102
Top	chr6	9200001	9300000	g14797.t1	PREDICTED: vesicular acetylcholine transporter	GO:0006635; GO:0006836; GO:0042908; GO:0055085; GO:0016746; GO:0042910; GO:0016021
Top	chr6	9300001	9400000	g14798.t1	RPE-retinal G protein-coupled receptor	GO:0007186; GO:0007601; GO:0007602; GO:0004930; GO:0009881; GO:0005887
Top	chr6	9300001	9400000	g14799.t1	Leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 1, partial	
Top	chr6	9300001	9400000	g14800.t1	Leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 1, partial	GO:0016021
Top	chr6	9300001	9400000	g14801.t1	leucine-rich repeat, immunoglobulin-like domain and transmembrane domain-containing protein 2	GO:0016021
Top	chr6	9400001	9500000	NA		
Top	chr6	9700001	9800000	g14803.t1	hypothetical protein N308_15335, partial	
Top	chr6	9900001	10000000	NA		
Top	chr6	10200001	10300000	NA		
Top	chr6	10300001	10400000	NA		
Top	chr6	17700001	17800000	g14920.t1	retinol-binding protein 3	GO:0006508; GO:0008236
Top	chr6	17700001	17800000	g14921.t1	growth/differentiation factor 2	GO:0007165; GO:0008083; GO:0005576
Top	chr6	17700001	17800000	g14922.t1	Bone morphogenetic protein 3B, partial	GO:0001649; GO:0007165; GO:0005125; GO:0008083; GO:0005615
Top	chr6	17900001	18000000	NA		

Top	chr6	18000001	18100000	g14924.t1	Mitogen-activated protein kinase 8	GO:0007258; GO:0009411; GO:0010628; GO:0016241; GO:0018105; GO:0018107; GO:0031063; GO:0031281; GO:0031343; GO:0032091; GO:0034198; GO:0034614; GO:0038095; GO:0042752; GO:0043065; GO:0043066; GO:0051090; GO:0051247; GO:0071222; GO:0071260; GO:0071276; GO:0090045; GO:0090398; GO:1900740; GO:1902595; GO:0004705; GO:0004712; GO:0005524; GO:0019903; GO:0035033; GO:0042826; GO:0106310; GO:0120283; GO:0005654; GO:0005739; GO:0005829; GO:0030424; GO:0045202; GO:0097441
Top	chr6	18000001	18100000	g14925.t1	rho GTPase-activating protein 22 isoform X4	GO:0007165; GO:0050790; GO:0099175; GO:0005096; GO:0098978
Top	chr6	20200001	20300000	g14977.t1	PREDICTED: cytochrome P450 26A1 isoform X2	GO:0006805; GO:0034653; GO:0001972; GO:0005506; GO:0008401; GO:0016709; GO:0020037; GO:0062183
Top	chr7	11500001	11600000	NA		
Top	chr7	11600001	11700000	g14209.t1	PREDICTED: enteropeptidase	
Top	chr7	13600001	13700000	g14224.t1	transcription factor CP2-like protein 1	GO:0000122; GO:0000902; GO:0002070; GO:0007028; GO:0007431; GO:0008340; GO:0045927; GO:0000978; GO:0003700; GO:0005634; GO:0005737; GO:0016020
Top	chr7	18300001	18400000	g14304.t1	PREDICTED: sodium channel protein type 3 subunit alpha isoform X2	GO:0034765; GO:0035725; GO:0005244; GO:0005248; GO:0001518
Top	chr7	23400001	23500000	NA		
Top	chr8	5800001	5900000	g15370.t1	PREDICTED: UPF0587 protein C1orf123 homolog isoform X1	GO:0008380; GO:0008270; GO:0005681; GO:0035145
Top	chr8	5800001	5900000	g15371.t1	carnitine O-palmitoyltransferase 2, mitochondrial	GO:0001676; GO:0006635; GO:0009437; GO:0120162; GO:0004095; GO:0005654; GO:0005730; GO:0005739
Top	chr8	5800001	5900000	g15372.t1	PREDICTED: excitatory amino acid transporter 5 isoform X2	GO:0015807; GO:0098656; GO:1902475; GO:0015293; GO:0016021

Top	chr8	13900001	14000000	NA		
Top	chr8	15900001	16000000	g15575.t1	Leucine-rich repeat-containing protein 8B	GO:0098656; GO:0005886; GO:0016021
Top	chr8	15900001	16000000	g15576.t1	hypothetical protein	
Top	chr10	20100001	20200000	NA		
Top	chr10	20200001	20300000	NA		
Top	chr11	21000001	21100000	g17270.t1	gliomedin	GO:0045162
Top	chr11	21000001	21100000	g17271.t1	aromatase	GO:0004497; GO:0005506; GO:0016705; GO:0020037; GO:0016021
Top	chr11	21100001	21200000	g17272.t1	ralBP1-associated Eps domain-containing protein 2	GO:0006278; GO:0003964
Top	chr11	21100001	21200000	g17273.t1	hypothetical protein llap_9627	
Top	chr11	21100001	21200000	g17274.t1	PREDICTED: tumor necrosis factor alpha-induced protein 8-like protein 3	GO:0015914; GO:0042981; GO:0043552; GO:0048017; GO:0051897; GO:0070374; GO:0120009; GO:0008526; GO:0035091; GO:0005654; GO:0005829; GO:0005886
Top	chr13	21300001	21400000	NA		
Top	chr18	100001	200000	g20088.t1	ras-specific guanine nucleotide-releasing factor RalGPS1 isoform X2	GO:0007264; GO:0050790; GO:0005085
Top	chr18	100001	200000	g20089.t1	angiopoietin-related protein 2	
Top	chr19	2900001	3000000	g19838.t1	PREDICTED: BUB3-interacting and GLEBS motif-containing protein ZNF207 isoform X5	
Top	chr19	2900001	3000000	g19839.t1	hypothetical protein H355_014863	GO:0006886; GO:0099041; GO:0005802; GO:0031410
Top	chr19	2900001	3000000	g19840.t1	rhomboid-related protein 3 isoform A	GO:0006508; GO:0004252; GO:0005509; GO:0016021
Top	chr24	3400001	3500000	g22433.t1	hypothetical protein	GO:0016021

Top	chr24	3400001	3500000	g22434.t1	NADPH-dependent aldo-keto reductase, chloroplastic	GO:0016491; GO:0016020; GO:0016021
Top	chrZ	7100001	7200000	g9573.t1	hypothetical protein CIB84_001467	GO:0006545; GO:0006729; GO:0031427; GO:0035999; GO:0046654; GO:0051000; GO:2000121; GO:0003729; GO:0004146; GO:0050661; GO:0005634; GO:0005739
Top	chrZ	9000001	9100000	NA		
Top	chrZ	9100001	9200000	NA		
Top	chrZ	9200001	9300000	g9633.t1	PREDICTED: LOW QUALITY PROTEIN: A-kinase anchor protein 2-like	GO:0008360; GO:0016310; GO:0016301; GO:0016020
Top	chrZ	9200001	9300000	g9634.t1	LOW QUALITY PROTEIN: thymic stromal cotransporter homolog	GO:0043029; GO:0045580; GO:0048538; GO:0055085; GO:0070233; GO:0070430; GO:0022857; GO:0009986; GO:0016021
Top	chrZ	9900001	10000000	g9656.t1	hypothetical protein H355_005386	GO:0000122; GO:0009314; GO:0033138; GO:0043388; GO:0045454; GO:0046826; GO:0071731; GO:1903206; GO:2000170; GO:0042803; GO:0047134; GO:0005576; GO:0005654; GO:0005829
Top	chrZ	10100001	10200000	g9658.t1	PREDICTED: muscle, skeletal receptor tyrosine-protein kinase isoform X3	GO:0008582; GO:0018108; GO:0030154; GO:0004712; GO:0004714; GO:0005524; GO:0016021; GO:0045211
Top	chrZ	10400001	10500000	g9663.t1	PREDICTED: atrial natriuretic peptide receptor 1-like isoform X8	GO:0006182; GO:0006468; GO:0007601; GO:0035556; GO:0004383; GO:0004672; GO:0005524
Top	chrZ	10400001	10500000	g9664.t1	PREDICTED: transmembrane protein 215	GO:0016021
Top	chrZ	10800001	10900000	NA		
Top	chrZ	11500001	11600000	g9699.t1	hypothetical protein C5E44_35305, partial	
Top	chrZ	11500001	11600000	g9700.t1	hypothetical protein llap_3480	GO:0007049; GO:0009790; GO:0010468; GO:0061780; GO:0003682; GO:0005634; GO:0016020; GO:0016021
Top	chrZ	14100001	14200000	g9780.t1	LOW QUALITY PROTEIN:	

					actin-like protein 7B	
Top	chrZ	14100001	14200000	g9781.t1	PREDICTED: aprataxin	GO:0006259; GO:0006974; GO:0005488; GO:0016787; GO:0031981
Top	chrZ	14100001	14200000	g9782.t1	PREDICTED: dnaJ homolog subfamily A member 1	GO:0006457; GO:0007283; GO:0009408; GO:0030317; GO:0030521; GO:0005524; GO:0030544; GO:0046872; GO:0050750; GO:0051082; GO:0016020
Top	chrZ	14100001	14200000	g9783.t1	PREDICTED: WD40 repeat-containing protein SMU1 isoform X1	GO:0000381; GO:0005737; GO:0016607; GO:0071005
Top	chrZ	14100001	14200000	g9784.t1	PREDICTED: beta-1,4-galactosyltransferase 1	GO:0005975; GO:0006486; GO:0016757; GO:0016021; GO:0032580
Top	chrZ	14400001	14500000	g9794.t1	glycerol kinase	GO:0003824
Top	chrZ	22100001	22200000	g10010.t1	tyrosine-protein kinase Fer	GO:0008283; GO:0018108; GO:0030335; GO:0031532; GO:0042058; GO:0051092; GO:0004713; GO:0005154
Top	chrZ	22100001	22200000	g10011.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Top	chrZ	22100001	22200000	g10012.t1	PREDICTED: WD repeat-containing protein 70, partial	GO:0016787
Top	chrZ	24600001	24700000	g10104.t1	PREDICTED: gypsy retrotransposon integrase-like protein 1	GO:0015074; GO:0003676
Top	chrZ	31900001	32000000	g10298.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	37400001	37500000	g10478.t1	hypothetical protein FNYG_14083	GO:0005975; GO:0006401; GO:0034427; GO:0016740; GO:0016787; GO:0055087
Top	chrZ	37400001	37500000	g10479.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016020; GO:0016021

Top	chrZ	37400001	37500000	g10480.t1	hypothetical protein llap_10261	
Top	chrZ	38500001	38600000	g10502.t1	Putative olfactory receptor 7A2	GO:0006468; GO:0007155; GO:0007186; GO:0050911; GO:0055085; GO:0004672; GO:0004930; GO:0004984; GO:0005524; GO:0005540; GO:0022857; GO:0005576; GO:0016021
Top	chrZ	38500001	38600000	g10503.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016020; GO:0016021
Top	chrZ	38700001	38800000	g10510.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0003824
Top	chrZ	38700001	38800000	g10511.t1	hypothetical protein N300_05636, partial	
Top	chrZ	41600001	41700000	g10599.t1	PREDICTED: transient receptor potential cation channel subfamily M member 3-like, partial	GO:0016048; GO:0050951; GO:0051262; GO:0070588; GO:0005227; GO:0005262; GO:0005887
Top	chrZ	41600001	41700000	g10600.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Top	chrZ	41600001	41700000	g10601.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	41600001	41700000	g10602.t1	testosterone 17-beta-dehydrogenase 3-like	GO:0006278; GO:0003964
Top	chrZ	46500001	46600000	g10744.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	46500001	46600000	g10745.t1	hypothetical protein llap_4920	GO:0006278; GO:0003964

Top	chrZ	46500001	46600000	g10746.t1	hypothetical protein ASZ78_001127	GO:0006629
Top	chrZ	49300001	49400000	g10889.t1	hypothetical protein	
Top	chrZ	49300001	49400000	g10890.t1	hypothetical protein, partial	
Top	chrZ	49300001	49400000	g10891.t1	hypothetical protein, partial	
Top	chrZ	49400001	49500000	g10892.t1	PREDICTED: uncharacterized protein LOC107055438	GO:0006278; GO:0003964
Top	chrZ	49400001	49500000	g10893.t1	pol- hypothetical protein	GO:0006278; GO:0003964
Top	chrZ	49400001	49500000	g10894.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	50100001	50200000	g10926.t1	hypothetical protein N340_13660, partial	
Top	chrZ	50100001	50200000	g10927.t1	hypothetical protein llap_1659	GO:0006278; GO:0003964
Top	chrZ	50100001	50200000	g10928.t1	hypothetical protein	GO:0006278; GO:0003964
Top	chrZ	50100001	50200000	g10929.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	50100001	50200000	g10930.t1	PREDICTED: sodium/potassium/calcium exchanger 4	GO:0006278; GO:0003964; GO:0016020
Top	chrZ	50200001	50300000	g10931.t1	hypothetical protein N308_04312, partial	
Top	chrZ	50200001	50300000	g10932.t1	lysine-specific demethylase 4C	
Top	chrZ	50400001	50500000	g10940.t1	hypothetical protein H355_010529	GO:0007049; GO:0016567; GO:0016740; GO:0046872; GO:0005634
Top	chrZ	50700001	50800000	g10951.t1	programmed cell death 1 ligand 1	GO:0006955; GO:0007166; GO:0031295; GO:0042102; GO:0042130; GO:0071222; GO:0009897; GO:0016021
Top	chrZ	52000001	52100000	g10999.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed	GO:0006278; GO:0003964

					DNA polymerase from mobile element jockey-like, partial	
Top	chrZ	52000001	52100000	g11000.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	52000001	52100000	g11001.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003824; GO:0003964
Top	chrZ	52200001	52300000	g11004.t1	doublesex- and mab-3-related transcription factor 2	GO:0014807; GO:0045944; GO:0048706; GO:2000287; GO:0000977; GO:0001228; GO:0042802; GO:0046872; GO:0005634
Top	chrZ	53000001	53100000	g11025.t1	PREDICTED: pentatricopeptide repeat-containing protein 2, mitochondrial isoform X2	GO:0050684; GO:0005739
Top	chrZ	53000001	53100000	g11026.t1	hypothetical protein	GO:0006278; GO:0003964
Top	chrZ	53000001	53100000	g11027.t1	hypothetical protein H355_010108	GO:0006355; GO:0010824; GO:0003677; GO:0003700; GO:0016787
Top	chrZ	53300001	53400000	g11033.t1	hypothetical protein H355_012376, partial	GO:0072583; GO:0005905
Top	chrZ	56400001	56500000	g11111.t1	PREDICTED: zinc transporter 5	GO:0006882; GO:0071577; GO:0005385; GO:0005794; GO:0016021
Top	chrZ	56700001	56800000	g11116.t1	hypothetical protein AS28_10536, partial	
Top	chrZ	59000001	59100000	g11174.t1	hypothetical protein llap_7685	
Top	chrZ	61200001	61300000	NA		
Top	chrZ	62000001	62100000	g11254.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Top	chrZ	62000001	62100000	g11255.t1	hypothetical protein N308_04931, partial	

Top	chrZ	75000001	75100000	g11598.t1	Myelin protein regulatory factor	GO:0006355; GO:0006508; GO:0003677; GO:0003700; GO:0004222; GO:0008270
Top	chrZ	75000001	75100000	g11599.t1	hypothetical protein C8258_32110, partial	GO:0006278; GO:0003964
Top	chrZ	75000001	75100000	g11600.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Top	chrZ	75000001	75100000	g11601.t1	ubiquitin carboxyl-terminal hydrolase 4	GO:0016787
Top	chrZ	75000001	75100000	g11602.t1	PREDICTED: uncharacterized protein LOC104832909	GO:0090502; GO:0003676; GO:0004523
Top	chrZ	75000001	75100000	g11603.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	11800001	11900000	g296.t1	hypothetical protein AV530_012697	
Bottom	chr1	11800001	11900000	g297.t1	PREDICTED: inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	GO:0016020; GO:0016021
Bottom	chr1	11800001	11900000	g298.t1	PREDICTED: enteropeptidase	
Bottom	chr1	11800001	11900000	g299.t1	hypothetical protein llap_8938	GO:0006278; GO:0003964
Bottom	chr1	11900001	12000000	g300.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr1	11900001	12000000	g301.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	11900001	12000000	g302.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	11900001	12000000	g303.t1	reverse hypothetical protein	GO:0006278; GO:0003964

Bottom	chr1	12000001	12100000	g304.t1	PREDICTED: inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	GO:0016020; GO:0016021
Bottom	chr1	12000001	12100000	g305.t1	hypothetical protein N301_07267, partial	
Bottom	chr1	12100001	12200000	g306.t1	PREDICTED: inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	GO:0016020; GO:0016021
Bottom	chr1	12100001	12200000	g307.t1	PREDICTED: thyroid peroxidase-like, partial	GO:0003824
Bottom	chr1	12100001	12200000	g308.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Bottom	chr1	12100001	12200000	g309.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Bottom	chr1	31600001	31700000	g701.t1	DNA helicase restriction enzyme Type III R subunit	GO:0006306; GO:0003677; GO:0004386; GO:0005524; GO:0008170; GO:0016787
Bottom	chr1	31600001	31700000	g702.t1	hypothetical protein llap_3564	GO:0000973; GO:0006355; GO:0006406; GO:0006606; GO:0006796; GO:0015031; GO:0051028; GO:0000287; GO:0003824; GO:0004427; GO:0016787; GO:0017056; GO:0046872; GO:0005634; GO:0005635; GO:0005643; GO:0005737; GO:0016020; GO:0016021; GO:0031080; GO:0031965
Bottom	chr1	31600001	31700000	g703.t1	hypothetical protein N338_02566, partial	
Bottom	chr1	31800001	31900000	g707.t1	pol- hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	31800001	31900000	g708.t1	unknown	GO:0006278; GO:0010951; GO:0003824; GO:0003964; GO:0004867; GO:0005615
Bottom	chr1	31800001	31900000	g709.t1	hypothetical protein TREES_T100019790	GO:0007218
Bottom	chr1	31800001	31900000	g710.t1	rna-directed dna polymerase from mobile	GO:0006278; GO:0003964

					element jockey-like	
Bottom	chr1	31800001	31900000	g711.t1	hypothetical protein llap_15501	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Bottom	chr1	31800001	31900000	g712.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
Bottom	chr1	31800001	31900000	g713.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Bottom	chr1	31800001	31900000	g714.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr1	31800001	31900000	g715.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr1	45600001	45700000	g1180.t1	PREDICTED: uncharacterized protein LOC106029472	GO:0006811; GO:0006813; GO:0071805; GO:0005267; GO:0016020; GO:0016021
Bottom	chr1	45600001	45700000	g1181.t1	vps10 domain-containing receptor 2	GO:0006278; GO:0003964
Bottom	chr1	45600001	45700000	g1182.t1	hypothetical protein AV530_006604	GO:0015074; GO:0044260; GO:0003676; GO:0008270; GO:0046872; GO:0016020; GO:0016021
Bottom	chr1	45600001	45700000	g1183.t1	uncharacterized protein LOC113840877	GO:0016032; GO:0003676; GO:0008270; GO:0016787; GO:0046872
Bottom	chr1	45600001	45700000	g1184.t1	endogenous retrovirus group K member 18 Pol protein-like	GO:0006278; GO:0015074; GO:0090502; GO:0003677; GO:0003964; GO:0004523; GO:0008233; GO:0008270
Bottom	chr1	45600001	45700000	g1185.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC104366128	
Bottom	chr1	45600001	45700000	g1186.t1	hypothetical protein	GO:0006278; GO:0003964

					llap_2558	
Bottom	chr1	45600001	45700000	g1187.t1	hypothetical protein llap_1482	
Bottom	chr1	55400001	55500000	g1492.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	55400001	55500000	g1493.t1	-	
Bottom	chr1	55400001	55500000	g1494.t1	hypothetical protein llap_11890	GO:0006278; GO:0006468; GO:0016310; GO:0048544; GO:0000166; GO:0003964; GO:0004672; GO:0004674; GO:0004712; GO:0005524; GO:0016301; GO:0016740; GO:0106310; GO:0016020; GO:0016021
Bottom	chr1	55400001	55500000	g1495.t1	hypothetical protein DUI87_22405	GO:0003824
Bottom	chr1	55400001	55500000	g1496.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	55400001	55500000	g1497.t1	hypothetical protein	GO:0110165
Bottom	chr1	80500001	80600000	g2023.t1	hypothetical protein N329_07954, partial	GO:0006278; GO:0003964
Bottom	chr1	80500001	80600000	g2024.t1	PREDICTED: uncharacterized protein CXorf22	GO:0031514
Bottom	chr1	83800001	83900000	g2088.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	83800001	83900000	g2089.t1	-	
Bottom	chr1	83800001	83900000	g2090.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964

Bottom	chr1	83800001	83900000	g2091.t1	hypothetical protein llap_9712	GO:0006278; GO:0003964
Bottom	chr1	83800001	83900000	g2092.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr1	92400001	92500000	g2277.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	92400001	92500000	g2278.t1	hypothetical protein N334_14226, partial	
Bottom	chr1	95400001	95500000	g2339.t1	hypothetical protein llap_5614	
Bottom	chr1	95400001	95500000	g2340.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC109365085	GO:0019058; GO:0051701; GO:0016020; GO:0033644
Bottom	chr1	95400001	95500000	g2341.t1	Glutamate receptor 3	
Bottom	chr1	116700001	116800000	g2658.t1	hypothetical protein H355_014663	GO:0010951; GO:0002020; GO:0004869; GO:0005615; GO:0005730; GO:0005829
Bottom	chr1	116700001	116800000	g2659.t1	hypothetical protein H355_014666, partial	GO:0005758
Bottom	chr1	116700001	116800000	g2660.t1	importin subunit alpha-5	GO:0006607; GO:0014841; GO:0014901; GO:0042981; GO:0060828; GO:0099527; GO:0008139; GO:0061608; GO:0005654; GO:0005829; GO:0014069; GO:0030425; GO:0042564; GO:0098978
Bottom	chr1	116700001	116800000	g2661.t1	F-box only protein 40	GO:0002376; GO:0016567; GO:0035725; GO:0008270; GO:0015280; GO:0061630; GO:0016021
Bottom	chr1	117300001	117400000	g2702.t1	olfactory receptor 10AC1-like	GO:0007186; GO:0050911; GO:0004930; GO:0004984; GO:0005886; GO:0016021
Bottom	chr1	117300001	117400000	g2703.t1	ephrin type-A receptor 1 isoform X2	GO:0018108; GO:0048013; GO:0004712; GO:0005003; GO:0005524; GO:0005887
Bottom	chr1	117700001	117800000	g2717.t1	hypothetical protein llap_5660	
Bottom	chr1	117700001	117800000	g2718.t1	hypothetical protein N300_14845, partial	
Bottom	chr1	117700001	117800000	g2719.t1	hypothetical protein llap_6758	
Bottom	chr1	117700001	117800000	g2720.t1	Transient receptor potential	GO:0098703; GO:0005262; GO:0005887

					cation channel subfamily V member 5, partial	
Bottom	chr1	117700001	117800000	g2721.t1	hypothetical protein Z169_01071, partial	
Bottom	chr1	117700001	117800000	g2722.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0098703; GO:0003964; GO:0005262; GO:0005887
Bottom	chr1	117800001	117900000	g2723.t1	ephrin type-B receptor 5	GO:0001806; GO:0018108; GO:0032092; GO:0048013; GO:0050798; GO:2000525; GO:0004712; GO:0005003; GO:0005524; GO:0005887; GO:0009986
Bottom	chr1	117800001	117900000	g2724.t1	T-cell receptor beta chain V domain 2, partial	
Bottom	chr1	117900001	118000000	g2726.t1	trypsin-II	GO:0006508; GO:0007586; GO:0004252; GO:0046872; GO:0005615
Bottom	chr1	117900001	118000000	g2727.t1	trypsin I-P1 precursor	GO:0006508; GO:0007586; GO:0004252; GO:0046872; GO:0005615
Bottom	chr1	117900001	118000000	g2728.t1	T-cell receptor beta chain V domain, partial	
Bottom	chr1	117900001	118000000	g2729.t1	hypothetical protein Z169_01180, partial	
Bottom	chr1	117900001	118000000	g2730.t1	glycerol kinase	GO:0044237; GO:0016772
Bottom	chr1	117900001	118000000	g2731.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	117900001	118000000	g2732.t1	hypothetical protein	
Bottom	chr1	117900001	118000000	g2733.t1	-	
Bottom	chr1	117900001	118000000	g2734.t1	PREDICTED: neurofilament medium polypeptide-like	
Bottom	chr1	117900001	118000000	g2735.t1	-	
Bottom	chr1	117900001	118000000	g2736.t1	-	
Bottom	chr1	120300001	120400000	g2838.t1	-	
Bottom	chr1	120300001	120400000	g2839.t1	hypothetical protein	

					N300_11097, partial	
Bottom	chr1	120300001	120400000	g2840.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003676; GO:0003964
Bottom	chr1	120300001	120400000	g2841.t1	hypothetical protein BCIN_18g00150	GO:0006357; GO:0006397; GO:0006508; GO:0016567; GO:0016575; GO:0045892; GO:0055085; GO:0003677; GO:0003714; GO:0004198; GO:0008270; GO:0061630; GO:0000118; GO:0005737; GO:0016021
Bottom	chr1	120400001	120500000	g2842.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003723; GO:0003724; GO:0003964; GO:0005524
Bottom	chr1	120400001	120500000	g2843.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	120400001	120500000	g2844.t1	hypothetical protein N300_07752, partial	
Bottom	chr1	121000001	121100000	g2859.t1	rho guanine nucleotide exchange factor 5 isoform X1	GO:0035556; GO:0050790; GO:0005085; GO:0005737
Bottom	chr1	121000001	121100000	g2860.t1	zinc finger protein 239-like	
Bottom	chr1	123100001	123200000	g2901.t1	uncharacterized protein LOC112089161	
Bottom	chr1	123100001	123200000	g2902.t1	adaptin ear-binding coat-associated protein 1	GO:0006278; GO:0003964
Bottom	chr1	123100001	123200000	g2903.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Bottom	chr1	123100001	123200000	g2904.t1	hypothetical protein, partial	GO:0006278; GO:0007166; GO:0007186; GO:0003964; GO:0004930; GO:0016021
Bottom	chr1	124900001	125000000	g2945.t1	hypothetical protein C5E45_35180, partial	GO:0110165
Bottom	chr1	124900001	125000000	g2946.t1	PREDICTED:	GO:0006790; GO:0008146; GO:0042802; GO:0005739;

					sulfotransferase 4A1	GO:0005829
Bottom	chr1	125100001	125200000	g2952.t1	gamma-parvin isoform X1	GO:0007163; GO:0030031; GO:0031532; GO:0034446; GO:0003779; GO:0005737; GO:0005925; GO:0015629
Bottom	chr1	125100001	125200000	g2953.t1	PREDICTED: endogenous retrovirus group K member 113 Pol protein-like, partial	GO:0044260; GO:0090304; GO:0005488; GO:0016787
Bottom	chr1	125300001	125400000	g2955.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	125300001	125400000	g2956.t1	hypothetical protein Z169_01071, partial	
Bottom	chr1	125300001	125400000	g2957.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr1	125300001	125400000	g2958.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	125300001	125400000	g2959.t1	PREDICTED: mitochondrial fission process protein 1	GO:0009987; GO:0044238; GO:0071704; GO:0003824
Bottom	chr1	125300001	125400000	g2960.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	125400001	125500000	g2961.t1	PREDICTED: enteropeptidase	
Bottom	chr1	125400001	125500000	g2962.t1	hypothetical protein llap_7334	
Bottom	chr1	125400001	125500000	g2963.t1	hypothetical protein llap_8938	GO:0006278; GO:0003964
Bottom	chr1	125400001	125500000	g2964.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr1	125400001	125500000	g2965.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	125400001	125500000	g2966.t1	PREDICTED: DEP domain-containing protein 1B, partial	GO:0006278; GO:0003964
Bottom	chr1	125500001	125600000	g2967.t1	adaptin ear-binding coat-	GO:0003824

					associated protein 1	
Bottom	chr1	125500001	125600000	g2968.t1	hypothetical protein N301_07267, partial	
Bottom	chr1	125500001	125600000	g2969.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	125800001	125900000	g2976.t1	PREDICTED: thyroid peroxidase-like, partial	GO:0003824
Bottom	chr1	125800001	125900000	g2977.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Bottom	chr1	125800001	125900000	g2978.t1	PREDICTED: uncharacterized protein LOC104832910, partial	
Bottom	chr1	125800001	125900000	g2979.t1	PREDICTED: uncharacterized protein LOC106047159	GO:0016787
Bottom	chr1	125800001	125900000	g2980.t1	hypothetical protein llap_7334	
Bottom	chr1	127300001	127400000	g3014.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
Bottom	chr1	127300001	127400000	g3015.t1	hypothetical protein N301_07267, partial	
Bottom	chr1	127300001	127400000	g3016.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr1	127300001	127400000	g3017.t1	PREDICTED: inositol 1,4,5-trisphosphate receptor-interacting protein-like 1	GO:0016020; GO:0016021
Bottom	chr1	127300001	127400000	g3018.t1	PREDICTED: thyroid peroxidase-like, partial	GO:0003824
Bottom	chr1	127300001	127400000	g3019.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Bottom	chr1	127300001	127400000	g3020.t1	PREDICTED: DEP domain-	GO:0006278; GO:0003964

					containing protein 1B, partial	
Bottom	chr1	127300001	127400000	g3021.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
Bottom	chr1	127300001	127400000	g3022.t1	hypothetical protein N301_07267, partial	
Bottom	chr1	130000001	130100000	g3090.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0003964
Bottom	chr1	130000001	130100000	g3091.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr1	130000001	130100000	g3092.t1	hypothetical protein N308_12456, partial	
Bottom	chr1	139200001	139300000	g3287.t1	pro-MCH	GO:0007165; GO:0007268; GO:0030354; GO:0045202
Bottom	chr1	144800001	144900000	NA		
Bottom	chr1	146100001	146200000	g3453.t1	hypothetical protein C5F51_36525	GO:0016772; GO:0110165
Bottom	chr1	146100001	146200000	g3454.t1	Very low-density lipoprotein receptor, partial	GO:0006869; GO:0006897; GO:0008203; GO:0005509; GO:0030229; GO:0005905; GO:0016021; GO:0034361
Bottom	chr1	146100001	146200000	g3455.t1	hypothetical protein llap_2921	
Bottom	chr1	146100001	146200000	g3456.t1	hypothetical protein C5E46_35725	GO:0003824
Bottom	chr1	146100001	146200000	g3457.t1	hypothetical protein llap_19670	
Bottom	chr1	146100001	146200000	g3458.t1	PREDICTED: uncharacterized protein LOC105414956	GO:0015074; GO:0090502; GO:0003676; GO:0004523; GO:0046872
Bottom	chr1	146100001	146200000	g3459.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Bottom	chr1	146100001	146200000	g3460.t1	hypothetical protein	GO:0006278; GO:0003964

					DUI87_13280	
Bottom	chr1	162400001	162500000	g3811.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	162400001	162500000	g3812.t1	hypothetical protein Z169_01180, partial	
Bottom	chr1	162400001	162500000	g3813.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr1	162400001	162500000	g3814.t1	-	
Bottom	chr1	162500001	162600000	g3816.t1	hypothetical protein C8258_32110, partial	GO:0006278; GO:0003964
Bottom	chr1	162500001	162600000	g3817.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
Bottom	chr1	162500001	162600000	g3818.t1	hypothetical protein AS28_03768, partial	GO:0006278; GO:0003964
Bottom	chr1	162500001	162600000	g3819.t1	hypothetical protein AS27_13285, partial	
Bottom	chr1	162500001	162600000	g3820.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	170500001	170600000	g3934.t1	ankyrin repeat, SAM and basic leucine zipper domain-containing protein 1-like	GO:0009987; GO:0022414; GO:0110165
Bottom	chr1	170500001	170600000	g3935.t1	ankyrin repeat, SAM and basic leucine zipper domain-containing protein 1 isoform X4	GO:0007283; GO:0030154; GO:0031047; GO:0051321; GO:0005737
Bottom	chr1	171200001	171300000	g3951.t1	hypothetical protein llap_16894	
Bottom	chr1	171200001	171300000	g3952.t1	rna-directed dna	GO:0006278; GO:0003964; GO:0016021

					polymerase from mobile element jockey-like	
Bottom	chr1	177600001	177700000	g4106.t1	hypothetical protein N305_08602, partial	GO:0006278; GO:0003964
Bottom	chr1	177600001	177700000	g4107.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964; GO:0004518; GO:0005488
Bottom	chr1	177600001	177700000	g4108.t1	hypothetical protein llap_3979	GO:0006278; GO:0003964
Bottom	chr1	177600001	177700000	g4109.t1	hypothetical protein HGM15179_012154	
Bottom	chr1	177600001	177700000	g4110.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Bottom	chr1	177600001	177700000	g4111.t1	hypothetical protein llap_3156	GO:0006278; GO:0003964
Bottom	chr1	177600001	177700000	g4112.t1	hypothetical protein Anapl_14099	
Bottom	chr1	177600001	177700000	g4113.t1	hypothetical protein N338_09572, partial	
Bottom	chr1	177700001	177800000	g4114.t1	hypothetical protein HGM15179_019012	
Bottom	chr1	177700001	177800000	g4115.t1	hypothetical protein llap_1028	
Bottom	chr1	177700001	177800000	g4116.t1	PREDICTED: mitochondrial fission process protein 1	GO:0003824
Bottom	chr1	177700001	177800000	g4117.t1	hypothetical protein C5E46_35725	GO:0006278; GO:0003964
Bottom	chr1	186600001	186700000	g4294.t1	hypothetical protein llap_9744	GO:0006811; GO:0006813; GO:0071805; GO:0090305; GO:0004519; GO:0005267; GO:0016020; GO:0016021
Bottom	chr1	186600001	186700000	g4295.t1	hypothetical protein DUI87_13214	GO:0030215; GO:0005576

Bottom	chr1	186600001	186700000	g4296.t1	uncharacterized protein GVQW3	GO:0032259; GO:0003676; GO:0008168; GO:0016740
Bottom	chr1	186900001	187000000	NA		
Bottom	chr1	187200001	187300000	g4306.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr1	187200001	187300000	g4307.t1	hypothetical protein	
Bottom	chr2	14900001	15000000	g4727.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Bottom	chr2	14900001	15000000	g4728.t1	PREDICTED: CUB and sushi domain-containing protein 3-like	GO:0016021
Bottom	chr2	14900001	15000000	g4729.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	14900001	15000000	g4730.t1	hypothetical protein AS27_10993, partial	
Bottom	chr2	14900001	15000000	g4731.t1	glycerol kinase	GO:0006278; GO:0016310; GO:0003964; GO:0009374; GO:0016301; GO:0005576
Bottom	chr2	15800001	15900000	g4762.t1	hypothetical protein C5E45_35180, partial	GO:0009987
Bottom	chr2	15800001	15900000	g4763.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr2	15800001	15900000	g4764.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	15800001	15900000	g4765.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	15800001	15900000	g4766.t1	rna-directed dna	GO:0006278; GO:0003723; GO:0003964; GO:0005737

					polymerase from mobile element jockey-hypothetical protein	
Bottom	chr2	15800001	15900000	g4767.t1	-	
Bottom	chr2	29500001	29600000	g5057.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	29500001	29600000	g5058.t1	hypothetical protein llap_6967	
Bottom	chr2	29500001	29600000	g5059.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr2	29500001	29600000	g5060.t1	PREDICTED: mitochondrial fission process protein 1	GO:0003824
Bottom	chr2	29600001	29700000	g5061.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0009968; GO:0003964
Bottom	chr2	29600001	29700000	g5062.t1	hypothetical protein llap_6172	GO:0006278; GO:0003964
Bottom	chr2	29600001	29700000	g5063.t1	hypothetical protein N308_06564, partial	
Bottom	chr2	29600001	29700000	g5064.t1	hypothetical protein llap_3199	GO:0006278; GO:0003964; GO:0016021
Bottom	chr2	29700001	29800000	g5066.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr2	29700001	29800000	g5067.t1	RNA-directed DNA polymerase from mobile element jockey, partial	GO:0006278; GO:0003964
Bottom	chr2	29800001	29900000	g5068.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	29800001	29900000	g5069.t1	hypothetical protein N300_01730, partial	
Bottom	chr2	29800001	29900000	g5070.t1	hypothetical protein llap_7061	GO:0006278; GO:0009166; GO:0000166; GO:0003676; GO:0003824; GO:0003964; GO:0008270; GO:0016787

Bottom	chr2	29800001	29900000	g5071.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	39300001	39400000	g5252.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr2	39300001	39400000	g5253.t1	hypothetical protein llap_11827	GO:0006278; GO:0003964
Bottom	chr2	39300001	39400000	g5254.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	39400001	39500000	g5255.t1	hypothetical protein AV530_012697	
Bottom	chr2	39400001	39500000	g5256.t1	hypothetical protein N308_08963, partial	
Bottom	chr2	39400001	39500000	g5257.t1	hypothetical protein llap_13557	
Bottom	chr2	39500001	39600000	g5258.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0003964; GO:0016021
Bottom	chr2	39500001	39600000	g5259.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr2	43400001	43500000	g5319.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003676; GO:0003964; GO:0008270; GO:0016021
Bottom	chr2	43400001	43500000	g5320.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0003824
Bottom	chr2	43400001	43500000	g5321.t1	hypothetical protein N334_06004, partial	
Bottom	chr2	63000001	63100000	g5732.t1	hypothetical protein Anapl_18080, partial	GO:0016310; GO:0016301; GO:0019894; GO:0005871
Bottom	chr2	63000001	63100000	g5733.t1	hypothetical protein Anapl_04608	GO:0008643; GO:0005576; GO:0016020; GO:0016021
Bottom	chr2	63000001	63100000	g5734.t1	pol- hypothetical protein	GO:0006278; GO:0003964
Bottom	chr2	63000001	63100000	g5735.t1	hypothetical protein	GO:0006278; GO:0006325; GO:0006629; GO:0016043;

					HGM15179_000422	GO:0030036; GO:0032259; GO:0034968; GO:0003779; GO:0003964; GO:0008168; GO:0008270; GO:0016740; GO:0018024; GO:0031267; GO:0046872; GO:0000775; GO:0005634; GO:0016020; GO:0016021
Bottom	chr2	66500001	66600000	g5786.t1	RNA-directed DNA polymerase from mobile element jockey, partial	GO:0006278; GO:0003964
Bottom	chr2	66900001	67000000	g5796.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr2	66900001	67000000	g5797.t1	hypothetical protein N338_01387, partial	
Bottom	chr2	66900001	67000000	g5798.t1	LOW QUALITY PROTEIN: uncharacterized protein LOC116243389	GO:0006278; GO:0003964
Bottom	chr2	67000001	67100000	g5799.t1	hypothetical protein llap_7885	GO:0006278; GO:0003723; GO:0003964
Bottom	chr2	67000001	67100000	g5800.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr2	70400001	70500000	g5863.t1	Putative Ser/Thr protein kinase KinX	GO:0016310; GO:0016301
Bottom	chr2	72400001	72500000	NA		
Bottom	chr2	87200001	87300000	g6146.t1	hypothetical protein N301_07267, partial	
Bottom	chr2	87200001	87300000	g6147.t1	LOW QUALITY PROTEIN: uncharacterized protein LOC116243389	GO:0006278; GO:0003964
Bottom	chr2	87200001	87300000	g6148.t1	mitochondrial fission process protein 1	GO:0006278; GO:0010951; GO:0016032; GO:0003964; GO:0004867
Bottom	chr2	93300001	93400000	g6324.t1	leucine-rich repeat flightless-interacting protein 2	GO:0006355; GO:0030275
Bottom	chr2	93300001	93400000	g6325.t1	PREDICTED: uncharacterized protein	

					LOC104156850	
Bottom	chr2	93300001	93400000	g6326.t1	hypothetical protein llap_7996	GO:0006278; GO:0003964
Bottom	chr2	93400001	93500000	g6328.t1	uncharacterized protein LOC101802146	
Bottom	chr2	93400001	93500000	g6329.t1	T-cell receptor gamma chain V domain, partial	
Bottom	chr2	93400001	93500000	g6330.t1	hypothetical protein Anapl_17715	
Bottom	chr2	93400001	93500000	g6331.t1	T-cell receptor gamma chain V domain, partial	
Bottom	chr2	98500001	98600000	g6430.t1	hypothetical protein	
Bottom	chr2	123000001	123100000	NA		
Bottom	chr2	124600001	124700000	g6879.t1	hypothetical protein Y956_16064, partial	
Bottom	chr2	124600001	124700000	g6880.t1	reverse transcriptase family protein, partial	GO:0003824; GO:0016021
Bottom	chr2	124600001	124700000	g6881.t1	ubiquitin carboxyl-terminal hydrolase 4	GO:0016787
Bottom	chr3	1	100000	g7366.t1	hypothetical protein	
Bottom	chr3	1	100000	g7367.t1	hypothetical protein llap_19699	GO:0006278; GO:0022008; GO:0003824; GO:0003964
Bottom	chr3	1	100000	g7368.t1	hypothetical protein CIB84_008264, partial	GO:0055129; GO:0004350
Bottom	chr3	1	100000	g7369.t1	delta-1-pyrroline-5-carboxylate synthase isoform X5	GO:0055129; GO:0004350; GO:0050661
Bottom	chr3	1	100000	g7370.t1	ras and Rab interactor 2 isoform X2	GO:0007165; GO:0016192; GO:0050790; GO:0005085; GO:0005096
Bottom	chr3	45700001	45800000	g8129.t1	CAP-Gly domain-containing linker protein 1-like	
Bottom	chr3	45700001	45800000	g8130.t1	ankyrin repeat domain-containing protein 26-like	

					isoform X2	
Bottom	chr3	45700001	45800000	g8131.t1	reverse transcriptase family protein, partial	GO:0003824
Bottom	chr3	54700001	54800000	g8270.t1	hypothetical protein	
Bottom	chr3	54700001	54800000	g8271.t1	hypothetical protein C5E45_35180, partial	GO:0009987
Bottom	chr3	54700001	54800000	g8272.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	54700001	54800000	g8273.t1	TRAF2 and NCK-interacting protein kinase	GO:0006468; GO:0007155; GO:0055085; GO:0004672; GO:0005524; GO:0005540; GO:0022857; GO:0005576; GO:0016021
Bottom	chr3	54700001	54800000	g8274.t1	hypothetical protein N301_07267, partial	
Bottom	chr3	54800001	54900000	g8275.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0010951; GO:0003964; GO:0004867
Bottom	chr3	54800001	54900000	g8276.t1	adaptin ear-binding coat-associated protein 1	GO:0003824
Bottom	chr3	54800001	54900000	g8277.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	54800001	54900000	g8278.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	54900001	55000000	g8279.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	54900001	55000000	g8280.t1	hypothetical protein DUI87_05917	GO:0006278; GO:0003964
Bottom	chr3	54900001	55000000	g8281.t1	-	
Bottom	chr3	55000001	55100000	g8282.t1	rna-directed dna polymerase from mobile	GO:0006278; GO:0003964; GO:0005515; GO:0005794; GO:0016020

					element jockey-like	
Bottom	chr3	55000001	55100000	g8283.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	55000001	55100000	g8284.t1	nedd4-binding protein 2-like 2	GO:0003824
Bottom	chr3	66000001	66100000	g8476.t1	hypothetical protein llap_7371	
Bottom	chr3	66100001	66200000	g8477.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	66100001	66200000	g8478.t1	Kinesin-associated protein 3	GO:0019894; GO:0005871
Bottom	chr3	66100001	66200000	g8479.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	66100001	66200000	g8480.t1	hypothetical protein N301_12929, partial	
Bottom	chr3	75300001	75400000	NA		
Bottom	chr3	75700001	75800000	g8652.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr3	75700001	75800000	g8653.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr3	75700001	75800000	g8654.t1	hypothetical protein N300_08864, partial	
Bottom	chr3	86300001	86400000	g8848.t1	-	
Bottom	chr3	86300001	86400000	g8849.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0016020
Bottom	chr3	86500001	86600000	g8852.t1	adaptin ear-binding coat-	GO:0006278; GO:0003964

					associated protein 1	
Bottom	chr3	86600001	86700000	g8853.t1	hypothetical protein DUI87_30491	GO:0006278; GO:0090305; GO:0003676; GO:0003964; GO:0004519; GO:0046872
Bottom	chr3	86600001	86700000	g8854.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr3	86600001	86700000	g8855.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr3	86600001	86700000	g8856.t1	hypothetical protein AS27_13285, partial	
Bottom	chr3	86600001	86700000	g8857.t1	hypothetical protein	
Bottom	chr3	86700001	86800000	g8858.t1	reverse transcriptase family protein, partial	GO:0006278; GO:0003964
Bottom	chr3	86700001	86800000	g8859.t1	hypothetical protein llap_1917	
Bottom	chr3	86700001	86800000	g8860.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr3	86700001	86800000	g8861.t1	PREDICTED: enteropeptidase	GO:0006278; GO:0003964
Bottom	chr3	104400001	104500000	g9181.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	104400001	104500000	g9182.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr3	104400001	104500000	g9183.t1	titin isoform X3	GO:0016021
Bottom	chr3	106000001	106100000	g9221.t1	hypothetical protein	
Bottom	chr3	106000001	106100000	g9222.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr3	106000001	106100000	g9223.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr3	106000001	106100000	g9224.t1	hypothetical protein	GO:0006278; GO:0003964

Bottom	chr3	106800001	106900000	g9247.t1	Intraflagellar transport protein 88 like protein, partial	GO:0007165; GO:0007186; GO:0015031; GO:0090150; GO:0004930; GO:0004990; GO:0005000; GO:0032977; GO:0005886; GO:0016020; GO:0016021; GO:0031514
Bottom	chr4	9600001	9700000	g11867.t1	hypothetical protein DUI87_07850	
Bottom	chr4	9600001	9700000	g11868.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr4	13000001	13100000	g11918.t1	PREDICTED: LOW QUALITY PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr4	13000001	13100000	g11919.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC109279696, partial	GO:0006278; GO:0003964
Bottom	chr4	19200001	19300000	g12014.t1	---NA---	
Bottom	chr4	19200001	19300000	g12015.t1	CC171 protein	GO:0016032; GO:0003676; GO:0008270; GO:0046872
Bottom	chr4	19200001	19300000	g12016.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr4	19200001	19300000	g12017.t1	adaptin ear-binding coat-associated protein 1	GO:0009987; GO:0016829; GO:0043167; GO:0016020
Bottom	chr4	19200001	19300000	g12018.t1	glycerol kinase	GO:0016310; GO:0016301
Bottom	chr4	19400001	19500000	g12022.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr4	19400001	19500000	g12023.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003824; GO:0003964; GO:0005737; GO:0045202
Bottom	chr4	19400001	19500000	g12024.t1	adaptin ear-binding coat-associated protein 1	GO:0006278; GO:0003964

Bottom	chr4	19500001	19600000	g12025.t1	hypothetical protein N339_02630, partial	
Bottom	chr4	19600001	19700000	g12026.t1	hypothetical protein N334_12917, partial	
Bottom	chr4	19600001	19700000	g12027.t1	ribonuclease H	GO:0006278; GO:0006310; GO:0006508; GO:0015074; GO:0090502; GO:0003723; GO:0003887; GO:0003964; GO:0004190; GO:0004523; GO:0046872
Bottom	chr4	19600001	19700000	g12028.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr4	19700001	19800000	g12029.t1	palmitoyltransferase ZDHHC21 isoform X1	GO:0006278; GO:0003964
Bottom	chr4	19700001	19800000	g12030.t1	protocadherin-7 isoform X1	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr4	34600001	34700000	NA		
Bottom	chr4	46500001	46600000	g12427.t1	hypothetical protein Anapl_05974	
Bottom	chr4	50700001	50800000	g12475.t1	RAGP1 protein	GO:0007165; GO:0050790; GO:0005096
Bottom	chr4	55600001	55700000	g12547.t1	---NA---	
Bottom	chr4	55600001	55700000	g12548.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr4	57200001	57300000	g12567.t1	RNA-directed DNA polymerase from mobile element jockey	GO:0006278; GO:0003964
Bottom	chr4	57600001	57700000	NA		
Bottom	chr4	57900001	58000000	NA		
Bottom	chr5	11200001	11300000	g13066.t1	hypothetical protein llap_3156	GO:0006278; GO:0003964
Bottom	chr5	18800001	18900000	g13194.t1	hypothetical protein llap_5264	GO:0008360; GO:0051301; GO:0016020; GO:0016021
Bottom	chr5	18800001	18900000	g13195.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021

Bottom	chr5	18800001	18900000	g13196.t1	hypothetical protein, partial	GO:0006278; GO:0007165; GO:0003964; GO:0004888; GO:0016020
Bottom	chr5	18800001	18900000	g13197.t1	hypothetical protein N335_14313, partial	
Bottom	chr5	18800001	18900000	g13198.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr5	18900001	19000000	g13199.t1	hypothetical protein, partial	
Bottom	chr5	18900001	19000000	g13200.t1	hypothetical protein AS28_03768, partial	
Bottom	chr5	18900001	19000000	g13201.t1	hypothetical protein Anapl_17196	GO:0016032; GO:0003676; GO:0008270; GO:0046872
Bottom	chr5	18900001	19000000	g13202.t1	hypothetical protein	
Bottom	chr5	37500001	37600000	NA		
Bottom	chr5	37600001	37700000	NA		
Bottom	chr5	40600001	40700000	g13723.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr5	40600001	40700000	g13724.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr5	40900001	41000000	g13725.t1	hypothetical protein AV530_012697	GO:0006278; GO:0003964
Bottom	chr5	51800001	51900000	g13906.t1	transcription factor E2F8, partial	GO:0006357; GO:0000978; GO:0046872; GO:0005634; GO:0005667
Bottom	chr5	52700001	52800000	g13916.t1	PREDICTED: protein kinase C-binding protein NELL1-like, partial	GO:0016310; GO:0005509; GO:0016301
Bottom	chr5	61300001	61400000	g14028.t1	hypothetical protein llap_584	
Bottom	chr6	13200001	13300000	g14841.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr6	17400001	17500000	NA		
Bottom	chr6	18700001	18800000	NA		
Bottom	chr6	19200001	19300000	NA		

Bottom	chr7	20200001	20300000	g14339.t1	serine/threonine-protein kinase tousled-like 1 isoform X2	GO:0006468; GO:0004672; GO:0005524
Bottom	chr7	23800001	23900000	g14401.t1	Nucampholin-like protein, partial	GO:0003723
Bottom	chr7	23800001	23900000	g14402.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr7	23800001	23900000	g14403.t1	hypothetical protein llap_2718	
Bottom	chr7	23800001	23900000	g14404.t1	MULTISPECIES: NAD(P)H nitroreductase	GO:0006355; GO:0006412; GO:0006418; GO:0006431; GO:0000049; GO:0000166; GO:0003677; GO:0003700; GO:0003723; GO:0004812; GO:0004825; GO:0005524; GO:0016491; GO:0016874; GO:0043565; GO:0046872; GO:0005737
Bottom	chr8	12900001	13000000	g15534.t1	hypothetical protein, partial	GO:0016021
Bottom	chr8	13000001	13100000	g15535.t1	hypothetical protein	
Bottom	chr8	15500001	15600000	g15569.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr8	17400001	17500000	g15614.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	17400001	17500000	g15615.t1	PREDICTED: uncharacterized protein LOC107055427	GO:0003824
Bottom	chr8	17400001	17500000	g15616.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	17500001	17600000	g15618.t1	hypothetical protein AS27_01106, partial	
Bottom	chr8	17500001	17600000	g15619.t1	-	
Bottom	chr8	17600001	17700000	g15620.t1	hypothetical protein N323_03669, partial	

Bottom	chr8	17600001	17700000	g15621.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr8	17700001	17800000	g15623.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	17700001	17800000	g15624.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	17700001	17800000	g15625.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr8	17700001	17800000	g15626.t1	RNA-directed DNA polymerase from mobile element jockey, partial	GO:0006278; GO:0003964
Bottom	chr8	17800001	17900000	g15627.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr8	17800001	17900000	g15628.t1	unknown	GO:0046872
Bottom	chr8	17800001	17900000	g15629.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr8	17800001	17900000	g15630.t1	hypothetical protein, partial	
Bottom	chr8	17800001	17900000	g15631.t1	hypothetical protein	
Bottom	chr8	17800001	17900000	g15632.t1	PREDICTED: uncharacterized protein LOC106893458	GO:0003824
Bottom	chr8	17900001	18000000	g15634.t1	hypothetical protein llap_9744	GO:0000723; GO:0051276; GO:0003677; GO:0043047; GO:0071949; GO:0000781
Bottom	chr8	17900001	18000000	g15635.t1	hypothetical protein llap_2221	GO:0016310; GO:0016301
Bottom	chr8	17900001	18000000	g15636.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Bottom	chr8	18100001	18200000	g15640.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr8	18100001	18200000	g15641.t1	hypothetical protein	GO:0009987

					llap_5535	
Bottom	chr8	18100001	18200000	g15642.t1	hypothetical protein llap_7613	GO:0006278; GO:0003964; GO:0016020; GO:0016021
Bottom	chr8	18200001	18300000	g15643.t1	vps10 domain-containing receptor 2	GO:0006278; GO:0055085; GO:0003964; GO:0004386; GO:0022857; GO:0016020; GO:0016021
Bottom	chr8	18200001	18300000	g15644.t1	PREDICTED: DEP domain-containing protein 1B, partial	GO:0006278; GO:0003964
Bottom	chr8	18200001	18300000	g15645.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	18200001	18300000	g15646.t1	hypothetical protein llap_6172	GO:0006278; GO:0003964
Bottom	chr8	18300001	18400000	g15648.t1	PREDICTED: BMP/retinoic acid-inducible neural-specific protein 3-like	GO:0045666; GO:0045930
Bottom	chr8	18300001	18400000	g15649.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0016310; GO:0003824; GO:0003964; GO:0016301
Bottom	chr8	18300001	18400000	g15650.t1	hypothetical protein H355_007216	GO:0045666; GO:0045930; GO:0071300; GO:0005783; GO:0030425; GO:0043025
Bottom	chr8	18600001	18700000	g15657.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	18600001	18700000	g15658.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr8	18600001	18700000	g15659.t1	T-cell surface glycoprotein CD1b-3-like isoform X1	GO:0016021
Bottom	chr8	24100001	24200000	g15801.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr8	24100001	24200000	g15802.t1	complement factor H	
Bottom	chr8	24100001	24200000	g15803.t1	coagulation factor XIII B	

					chain	
Bottom	chr8	24100001	24200000	g15804.t1	abnormal spindle-like microcephaly-associated protein	GO:0007051; GO:0005516; GO:0005634; GO:0005737
Bottom	chr8	31100001	31200000	NA		
Bottom	chr9	1	100000	g15911.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr9	1	100000	g15912.t1	-	
Bottom	chr9	1	100000	g15913.t1	nuclear cap-binding protein subunit 2 2	GO:0045292; GO:0000339; GO:0005634; GO:0005846
Bottom	chr9	1	100000	g15914.t1	Carboxypeptidase B, partial	GO:0006508; GO:0004181; GO:0008270
Bottom	chr9	1	100000	g15915.t1	type-1 angiotensin II receptor	GO:0001822; GO:0006954; GO:0007266; GO:0010873; GO:0019229; GO:0019722; GO:0032430; GO:0046718; GO:0051482; GO:0060326; GO:0086097; GO:1903589; GO:0001596; GO:0004945; GO:0031711; GO:0046982; GO:0005886; GO:0016021
Bottom	chr9	900001	1000000	g15927.t1	hypothetical protein ASZ78_014762	GO:0017121; GO:0017128; GO:0005886
Bottom	chr9	900001	1000000	g15928.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr9	1200001	1300000	g15932.t1	PREDICTED: A disintegrin and metalloproteinase with thrombospondin motifs 1	GO:0003676; GO:0016787
Bottom	chr9	22200001	22300000	g16349.t1	hypothetical protein ASZ78_012335	GO:0005975; GO:0030246; GO:0090599; GO:0016021
Bottom	chr9	22200001	22300000	g16350.t1	ABC transporter substrate-binding protein	GO:0042597
Bottom	chr10	1	100000	g17707.t1	hypothetical protein N322_04680, partial	
Bottom	chr10	1	100000	g17708.t1	hypothetical protein llap_1111	GO:0006278; GO:0003964
Bottom	chr10	1	100000	g17709.t1	PREDICTED: RNA-directed DNA polymerase from	GO:0006278; GO:0003964

					mobile element jockey-like, partial	
Bottom	chr10	1	100000	g17710.t1	transcriptional regulator Kaiso	GO:0000122; GO:0000978; GO:0001227; GO:0008327; GO:0005654; GO:0005730; GO:0005829; GO:0005886
Bottom	chr10	1	100000	g17711.t1	hypothetical protein ASZ78_011461	GO:0009617; GO:0016021
Bottom	chr10	1	100000	g17712.t1	hypothetical protein ASZ78_011461	GO:0009617; GO:0016021
Bottom	chr10	1	100000	g17713.t1	protein ATP1B4	GO:0006355; GO:0006813; GO:0006814; GO:0005637; GO:0005890
Bottom	chr10	1	100000	g17714.t1	lysosome-associated membrane glycoprotein 2 isoform X2	GO:0009267; GO:0061740; GO:0097352; GO:1905146; GO:0005765; GO:0005886; GO:0031902; GO:0097637
Bottom	chr10	20300001	20400000	g18082.t1	Ribosome-binding protein 1	GO:0006869; GO:0006914; GO:0007030; GO:0007041; GO:0015031; GO:0032456; GO:0042147; GO:0048193; GO:0048854; GO:0000938; GO:0005730; GO:0005829; GO:0016020; GO:1990745
Bottom	chr10	20300001	20400000	g18083.t1	fibroin heavy chain-like	GO:0006357; GO:0006606; GO:0003690; GO:0003712; GO:0017056; GO:0005634; GO:0044613
Bottom	chr10	20300001	20400000	g18084.t1	vacuolar protein sorting-associated protein 51 homolog isoform X2	GO:0006869; GO:0006914; GO:0007030; GO:0007041; GO:0015031; GO:0032456; GO:0042147; GO:0048193; GO:0048854; GO:0000938; GO:0005730; GO:0005829; GO:0016020; GO:1990745
Bottom	chr11	13800001	13900000	g17157.t1	hypothetical protein H355_004128	GO:0006511; GO:0016579; GO:0004843; GO:0008270
Bottom	chr12	1	100000	g17275.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr12	1	100000	g17276.t1	LOW QUALITY PROTEIN: putative ATP-dependent RNA helicase TDRD12	
Bottom	chr12	1	100000	g17277.t1	regulator of G-protein signaling 9-binding protein	GO:0007186; GO:0009968; GO:0050908; GO:0001750; GO:0016021
Bottom	chr12	1	100000	g17278.t1	ankyrin repeat domain-containing protein 27	

					isoform X4	
Bottom	chr12	100001	200000	g17279.t1	hypothetical protein llap_13075	GO:0098609; GO:0046872; GO:0016020
Bottom	chr12	100001	200000	g17280.t1	PREDICTED: neural-cadherin, partial	GO:0098609; GO:0046872; GO:0016020
Bottom	chr12	100001	200000	g17281.t1	hypothetical protein CIB84_004083, partial	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr12	100001	200000	g17282.t1	PREDICTED: probable C-mannosyltransferase DPY19L3 isoform X3	GO:0016757; GO:0016021
Bottom	chr12	300001	400000	g17285.t1	hypothetical protein C5E45_35180, partial	GO:0009987; GO:0050896; GO:0003824
Bottom	chr12	300001	400000	g17286.t1	adaptin ear-binding coat-associated protein 1	GO:0110165
Bottom	chr12	300001	400000	g17287.t1	hypothetical protein	
Bottom	chr12	300001	400000	g17288.t1	hypothetical protein llap_2250	GO:0006278; GO:0003964
Bottom	chr12	400001	500000	g17289.t1	hypothetical protein llap_451	GO:0016020
Bottom	chr12	400001	500000	g17290.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr12	400001	500000	g17291.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr12	12000001	12100000	NA		
Bottom	chr12	12300001	12400000	g17541.t1	cadherin-11	GO:0007156; GO:0030199; GO:0005509; GO:0008013; GO:0046983; GO:0005886; GO:0005911; GO:0016021
Bottom	chr12	12600001	12700000	NA		
Bottom	chr12	12800001	12900000	g17543.t1	hypothetical protein AS27_13285, partial	GO:0006278; GO:0003964
Bottom	chr12	12900001	13000000	NA		
Bottom	chr12	13000001	13100000	g17544.t1	TIGR00159 family protein	
Bottom	chr12	13200001	13300000	NA		

Bottom	chr13	1	100000	g16428.t1	ovochymase-2	GO:0006508; GO:0004252
Bottom	chr13	1	100000	g16429.t1	DCC-interacting protein 13-alpha isoform X1	GO:0007049; GO:0042802; GO:0001726; GO:0005634; GO:0031901; GO:0045335
Bottom	chr13	1	100000	g16430.t1	hypothetical protein N306_15064, partial	
Bottom	chr13	1	100000	g16431.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr13	4200001	4300000	g16496.t1	probable E3 ubiquitin-protein ligase makorin-2, partial	GO:0016740; GO:0046872
Bottom	chr13	4200001	4300000	g16497.t1	PREDICTED: MKRN2 opposite strand protein	
Bottom	chr13	4200001	4300000	g16498.t1	PREDICTED: tRNA-splicing endonuclease subunit Sen2	GO:0006388; GO:0090502; GO:0000213; GO:0003676; GO:0016829; GO:0000214
Bottom	chr13	4200001	4300000	g16499.t1	PREDICTED: peroxisome proliferator-activated receptor gamma	GO:0006357; GO:0030522; GO:0048511; GO:2000112; GO:0004879; GO:0008270; GO:0043565; GO:0005634; GO:0005737
Bottom	chr14	2000001	2100000	NA		
Bottom	chr14	2100001	2200000	g18129.t1	hypothetical protein ASZ78_014594	GO:0005829; GO:0045171; GO:0072686
Bottom	chr14	2100001	2200000	g18130.t1	hyaluronan mediated motility receptor	GO:0005540; GO:0005737; GO:0005856
Bottom	chr14	2100001	2200000	g18131.t1	PREDICTED: methionine adenosyltransferase 2 subunit beta	GO:0006556; GO:0006730; GO:0050790; GO:0016740; GO:0019899; GO:0048270; GO:0005634; GO:0048269
Bottom	chr14	7400001	7500000	g18201.t1	PREDICTED: protocadherin gamma-C3	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr14	7400001	7500000	g18202.t1	Protocadherin gamma-B5	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr14	7400001	7500000	g18203.t1	hypothetical protein CIB84_017199, partial	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr14	7400001	7500000	g18204.t1	LOW QUALITY PROTEIN:	GO:0007156; GO:0005509; GO:0005886; GO:0016021

					protocadherin gamma-B5-like	
Bottom	chr14	7400001	7500000	g18205.t1	PREDICTED: protocadherin beta-3-like	GO:0007156; GO:0005509; GO:0005886; GO:0016021
Bottom	chr15	1	100000	g18493.t1	hypothetical protein ASZ78_002272	GO:0016740
Bottom	chr15	1	100000	g18494.t1	ribosome biogenesis protein TSR3 homolog	GO:0000455; GO:0106388; GO:1904047; GO:0005737
Bottom	chr15	7300001	7400000	g18678.t1	ubiquitin carboxyl-terminal hydrolase 42	GO:0006511; GO:0016579; GO:0004843
Bottom	chr15	7300001	7400000	g18679.t1	PREDICTED: cytohesin-3 isoform X2	GO:0032012; GO:0050790; GO:0005085
Bottom	chr15	17000001	17100000	g18944.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr15	17000001	17100000	g18945.t1	LOW QUALITY PROTEIN: DNA repair endonuclease XPF	GO:0006281; GO:0090305; GO:0003697; GO:0004520; GO:0005634
Bottom	chr16	1	100000	g19334.t1	PREDICTED: lysine-specific demethylase 3A isoform X1	GO:0032259; GO:0033169; GO:0008168; GO:0016787; GO:0032454
Bottom	chr16	1	100000	g19335.t1	probable ATP-dependent RNA helicase DHX37	GO:0003676; GO:0004386; GO:0005524
Bottom	chr16	1	100000	g19336.t1	PREDICTED: BRI3-binding protein	GO:0005739; GO:0016021
Bottom	chr16	1	100000	g19337.t1	PREDICTED: acetoacetyl-CoA synthetase	GO:0006631; GO:0032024; GO:0005524; GO:0030729; GO:0005829
Bottom	chr16	100001	200000	g19338.t1	hypothetical protein llap_2130	
Bottom	chr16	100001	200000	g19339.t1	PREDICTED: transmembrane protein 132B	GO:0016021
Bottom	chr16	200001	300000	g19340.t1	transmembrane protein 132B isoform X1	GO:0016021
Bottom	chr16	200001	300000	g19341.t1	hypothetical protein	
Bottom	chr16	300001	400000	g19343.t1	hypothetical protein	

					AS28_03768, partial	
Bottom	chr16	300001	400000	g19344.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like, partial	GO:0006278; GO:0003964
Bottom	chr16	300001	400000	g19345.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chr16	300001	400000	g19346.t1	transmembrane protein 132B isoform X1	GO:0016021
Bottom	chr16	500001	600000	NA		
Bottom	chr17	100001	200000	g22772.t1	hypothetical protein N334_14226, partial	
Bottom	chr17	100001	200000	g22773.t1	C-type lectin domain family 9 member A-like isoform X1	GO:0016020; GO:0016021
Bottom	chr17	100001	200000	g22774.t1	LOW QUALITY PROTEIN: zinc finger protein 850	GO:0016021
Bottom	chr17	100001	200000	g22775.t1	LOW QUALITY PROTEIN: zinc finger protein 850	
Bottom	chr17	100001	200000	g22776.t1	zinc finger protein 883-like isoform X1	
Bottom	chr17	100001	200000	g22777.t1	LOW QUALITY PROTEIN: zinc finger protein 850	
Bottom	chr17	100001	200000	g22778.t1	butyrophilin subfamily 1 member A1-like	GO:0016020
Bottom	chr17	100001	200000	g22779.t1	butyrophilin subfamily 1 member A1-like, partial	GO:0016021
Bottom	chr17	100001	200000	g22780.t1	butyrophilin subfamily 1 member A1-like, partial	GO:0016021
Bottom	chr17	100001	200000	g22781.t1	butyrophilin subfamily 3 member A2	GO:0016021
Bottom	chr17	100001	200000	g22782.t1	butyrophilin subfamily 1 member A1, partial	GO:0016021

Bottom	chr17	100001	200000	g22783.t1	butyrophilin subfamily 3 member A3-like isoform X3	GO:0016021
Bottom	chr17	100001	200000	g22784.t1	butyrophilin subfamily 1 member A1 isoform X2	GO:0016021
Bottom	chr17	100001	200000	g22785.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr17	100001	200000	g22786.t1	zinc finger protein 883 isoform X1	GO:0006355; GO:0046872
Bottom	chr17	200001	300000	NA		
Bottom	chr18	1	100000	g20084.t1	PREDICTED: GTPase-activating Rap/Ran-GAP domain-like protein 3 isoform X6	GO:0065007
Bottom	chr18	1	100000	g20085.t1	PREDICTED: GTPase-activating Rap/Ran-GAP domain-like protein 3, partial	GO:0050790; GO:0051056; GO:0005096
Bottom	chr18	1	100000	g20086.t1	solute carrier family 2, facilitated glucose transporter member 8 isoform X1	GO:0008643; GO:0055085; GO:0022857; GO:0016021
Bottom	chr18	1	100000	g20087.t1	PREDICTED: F-box/WD repeat-containing protein 2 isoform X1	
Bottom	chr19	2600001	2700000	g19829.t1	importin subunit alpha-1	GO:0006606; GO:0061608; GO:0005737
Bottom	chr19	10800001	10900000	g20036.t1	PREDICTED: growth arrest-specific protein 7, partial	
Bottom	chr19	10800001	10900000	g20037.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr19	10800001	10900000	g20038.t1	hypothetical protein N340_09384, partial	
Bottom	chr19	10900001	11000000	g20040.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964

Bottom	chr19	11000001	11100000	g20042.t1	hypothetical protein N309_11457, partial	
Bottom	chr19	11000001	11100000	g20043.t1	unnamed protein product, partial	GO:0003774; GO:0003824; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11000001	11100000	g20044.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr19	11000001	11100000	g20045.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr19	11000001	11100000	g20046.t1	PREDICTED: myosin-3-like, partial	
Bottom	chr19	11000001	11100000	g20047.t1	PREDICTED: myosin-3-like, partial	GO:0003774; GO:0005524; GO:0051015; GO:0016459; GO:0030016
Bottom	chr19	11100001	11200000	g20049.t1	hypothetical protein AS28_00378, partial	GO:0006278; GO:0003964
Bottom	chr19	11100001	11200000	g20050.t1	myosin-1b-like	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11100001	11200000	g20051.t1	PREDICTED: myosin-1B-like, partial	
Bottom	chr19	11100001	11200000	g20052.t1	pol- hypothetical protein	GO:0006278; GO:0003964
Bottom	chr19	11100001	11200000	g20053.t1	hypothetical protein, partial	
Bottom	chr19	11100001	11200000	g20054.t1	hypothetical protein llap_19931	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11100001	11200000	g20055.t1	PREDICTED: myosin heavy chain, skeletal muscle-like, partial	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11200001	11300000	g20057.t1	PREDICTED: LOW QUALITY PROTEIN: myosin-3-like, partial	GO:0016459
Bottom	chr19	11200001	11300000	g20058.t1	hypothetical protein llap_8516	
Bottom	chr19	11200001	11300000	g20059.t1	PREDICTED: myosin heavy chain, skeletal muscle,	GO:0003774; GO:0005524; GO:0051015; GO:0016459

					adult-like	
Bottom	chr19	11200001	11300000	g20060.t1	MyHC1	GO:0003774; GO:0005516; GO:0005524; GO:0051015; GO:0016459; GO:0030016; GO:0032982
Bottom	chr19	11200001	11300000	g20061.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chr19	11200001	11300000	g20062.t1	myosin heavy chain, skeletal muscle, adult, partial	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11200001	11300000	g20063.t1	PREDICTED: myosin heavy chain, skeletal muscle, adult-like, partial	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11200001	11300000	g20064.t1	PREDICTED: myosin heavy chain, skeletal muscle, adult-like	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr19	11200001	11300000	g20065.t1	PREDICTED: myosin-3-like	GO:0003774; GO:0005524; GO:0051015; GO:0016459
Bottom	chr20	10800001	10900000	g20698.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr20	10800001	10900000	g20699.t1	1-acyl-sn-glycerol-3-phosphate acyltransferase alpha	GO:0008654; GO:0003841; GO:0016021
Bottom	chr20	10800001	10900000	g20700.t1	myeloperoxidase-like isoform X3	GO:0006979; GO:0098869; GO:0004601; GO:0020037
Bottom	chr20	10800001	10900000	g20701.t1	C-C motif chemokine 26 precursor	GO:0006955; GO:0007165; GO:0060326; GO:0008009; GO:0005615
Bottom	chr20	10800001	10900000	g20702.t1	uncharacterized protein LOC101791535	
Bottom	chr20	10800001	10900000	g20703.t1	LOW QUALITY PROTEIN: E3 ubiquitin-protein ligase TRIM39-like	GO:0008270
Bottom	chr20	10800001	10900000	g20704.t1	Uncharacterized protein C8orf45-like	GO:0032508; GO:0003677; GO:0005524
Bottom	chr20	11200001	11300000	g20719.t1	PREDICTED: LOW QUALITY	GO:0006278; GO:0003964

					PROTEIN: RNA-directed DNA polymerase from mobile element jockey-like, partial	
Bottom	chr20	11200001	11300000	g20720.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chr20	11200001	11300000	g20721.t1	PREDICTED: syntaxin-1A, partial	GO:0006836; GO:0006886; GO:0017157; GO:0061025; GO:0000149; GO:0005484; GO:0016021
Bottom	chr20	11200001	11300000	g20722.t1	probable 18S rRNA (guanine-N(7))-methyltransferase isoform X3	GO:0070476; GO:2000234; GO:0016435; GO:0046982; GO:0005654; GO:0005730; GO:0048471
Bottom	chr20	11200001	11300000	g20723.t1	PREDICTED: dnaJ homolog subfamily C member 30	GO:0016021
Bottom	chr20	11200001	11300000	g20724.t1	vacuolar protein sorting-associated protein 37D	GO:0015031; GO:0046907; GO:0070476; GO:0016435; GO:0000813; GO:0005634; GO:0031902
Bottom	chr20	11200001	11300000	g20725.t1	fibrinogen-like protein 1-like protein	GO:0031638; GO:0004252; GO:0005886
Bottom	chr20	11200001	11300000	g20726.t1	fibrinogen-like protein 1-like protein	
Bottom	chr20	11200001	11300000	g20727.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr21	10100001	10200000	g19209.t1	E3 ubiquitin-protein ligase Itchy homolog isoform X1	GO:0002669; GO:0032088; GO:0035519; GO:0043066; GO:0043161; GO:0045732; GO:0046329; GO:0046642; GO:0050687; GO:0051865; GO:0070534; GO:0070936; GO:0090085; GO:2000646; GO:0016874; GO:0043021; GO:0044389; GO:0045236; GO:0061630; GO:1990763; GO:0005654; GO:0005769; GO:0005886; GO:0005938; GO:0032991
Bottom	chr21	10100001	10200000	g19210.t1	PREDICTED: uncharacterized protein C20orf24 homolog	GO:0097250; GO:0016021
Bottom	chr22	1	100000	g21874.t1	LOW QUALITY PROTEIN:	GO:0006974; GO:0005634

					zinc finger and BTB domain-containing protein 40	
Bottom	chr22	1	100000	g21875.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0010951; GO:0003964; GO:0004867; GO:0005615
Bottom	chr22	1	100000	g21876.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr22	8000001	8100000	g22115.t1	matrix metalloproteinase-23 isoform X1	GO:0000003; GO:0006508; GO:0004222; GO:0008270; GO:0016021; GO:0031012
Bottom	chr24	1200001	1300000	g22377.t1	selenoprotein N isoform X1	
Bottom	chr24	1200001	1300000	g22378.t1	mitochondrial fission regulator 1-like	
Bottom	chr24	1200001	1300000	g22379.t1	hypothetical protein Anapl_17196	GO:0016032; GO:0003676; GO:0008270; GO:0046872
Bottom	chr24	1200001	1300000	g22380.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0006629; GO:0007165; GO:0003824; GO:0003964
Bottom	chr24	1500001	1600000	g22391.t1	PREDICTED: uncharacterized protein LOC106036278	GO:0006508; GO:0004190
Bottom	chr24	1500001	1600000	g22392.t1	reverse hypothetical protein	GO:0016021
Bottom	chr24	1500001	1600000	g22393.t1	PREDICTED: putative uncharacterized protein FLJ37770	GO:0003676; GO:0016740
Bottom	chr24	1500001	1600000	g22394.t1	PREDICTED: heterogeneous nuclear ribonucleoprotein R isoform X1	GO:0003729; GO:0005654; GO:0071013
Bottom	chr26	300001	400000	g21551.t1	PREDICTED: feather keratin 1-like isoform X1	GO:0007010; GO:0005200; GO:0005882

Bottom	chr26	300001	400000	g21552.t1	feather keratin 1-like	GO:0007010; GO:0005200; GO:0005882
Bottom	chr26	300001	400000	g21553.t1	feather keratin 1	GO:0007010; GO:0005200; GO:0005882
Bottom	chr26	300001	400000	g21554.t1	reverse hypothetical protein	GO:0016020
Bottom	chr26	300001	400000	g21555.t1	PREDICTED: feather beta keratin-like	GO:0007010; GO:0005200; GO:0005882
Bottom	chr26	300001	400000	g21556.t1	hypothetical protein AV530_009263	GO:0007010; GO:0005200; GO:0005882
Bottom	chr26	300001	400000	g21557.t1	LOW QUALITY PROTEIN: claw keratin-like	GO:0007010; GO:0005200; GO:0005882
Bottom	chr26	300001	400000	g21558.t1	keratin-associated protein 20-1-like	
Bottom	chr26	300001	400000	g21559.t1	keratin-associated protein 21-1-like isoform X1	GO:0016021
Bottom	chr27	6100001	6200000	NA		
Bottom	chr28	1	100000	g21246.t1	hypothetical protein CIB84_013086	GO:0035725; GO:0005272; GO:0016021
Bottom	chr28	1	100000	g21247.t1	hypothetical protein N321_09184, partial	GO:0006278; GO:0016310; GO:0003964; GO:0016301; GO:0016757
Bottom	chr28	1	100000	g21248.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr28	100001	200000	g21250.t1	endogenous retrovirus group K member 11 Pol protein-like	GO:0006278; GO:0015074; GO:0090502; GO:0003676; GO:0003964; GO:0004523; GO:0008270
Bottom	chr28	100001	200000	g21251.t1	PREDICTED: endogenous retrovirus group K member 18 Pol protein-like	GO:0006278; GO:0016032; GO:0003723; GO:0003887; GO:0003964
Bottom	chr28	100001	200000	g21252.t1	PREDICTED: uncharacterized protein LOC104839510	GO:0006278; GO:0016032; GO:0003723; GO:0003887; GO:0003964; GO:0016021
Bottom	chr28	100001	200000	g21253.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr28	100001	200000	g21254.t1	adaptin ear-binding coat-	GO:0006278; GO:0003964

					associated protein 1	
Bottom	chr28	100001	200000	g21255.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr28	200001	300000	g21256.t1	reverse hypothetical protein	GO:0006278; GO:0003964
Bottom	chr28	200001	300000	g21257.t1	feather keratin Cos2-3	GO:0007010; GO:0005200; GO:0005882
Bottom	chr28	200001	300000	g21258.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr28	200001	300000	g21259.t1	uncharacterized protein LOC110361212	GO:0003824
Bottom	chr28	200001	300000	g21260.t1	hypothetical protein AS27_11562, partial	
Bottom	chr28	200001	300000	g21261.t1	hypothetical protein llap_7802	
Bottom	chr28	300001	400000	g21263.t1	hypothetical protein ASZ78_010752	GO:0007010; GO:0005200; GO:0005882
Bottom	chr28	300001	400000	g21264.t1	hypothetical protein	GO:0006278; GO:0003964
Bottom	chr28	300001	400000	g21265.t1	T-cell receptor alpha chain V domain, partial	
Bottom	chr28	300001	400000	g21266.t1	adaptin ear-binding coat-associated protein 1	GO:0044237; GO:0016772
Bottom	chr28	300001	400000	g21267.t1	T-cell receptor alpha chain V domain 1, partial	GO:0042605
Bottom	chr28	300001	400000	g21268.t1	feather keratin Cos1-2-like	GO:0007010; GO:0005200; GO:0005882
Bottom	chr28	300001	400000	g21269.t1	hypothetical protein Z169_11379, partial	
Bottom	chr28	300001	400000	g21270.t1	hypothetical protein	GO:0006278; GO:0043154; GO:0003964; GO:0043027; GO:0016021
Bottom	chr28	400001	500000	g21272.t1	Gametocyte-specific factor 1, partial	GO:0007283; GO:0046872
Bottom	chr28	400001	500000	g21273.t1	PREDICTED: Golgi SNAP receptor complex member 2 isoform X1	GO:0015031; GO:0016192; GO:0061025; GO:0005484; GO:0005794; GO:0016021

Bottom	chr28	400001	500000	g21274.t1	PREDICTED: protein Wnt-9b	GO:0016055; GO:0005102; GO:0005576
Bottom	chr28	1700001	1800000	g21324.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr28	1700001	1800000	g21325.t1	hypothetical protein llap_3311	
Bottom	chr28	1700001	1800000	g21326.t1	PREDICTED: von Willebrand factor A domain-containing protein 5A-like, partial	
Bottom	chr28	1700001	1800000	g21327.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0016310; GO:0003824; GO:0003964; GO:0009374; GO:0016301; GO:0005576
Bottom	chr28	1800001	1900000	g21329.t1	PREDICTED: dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	GO:0006487; GO:0031647; GO:0043066; GO:0050790; GO:0008047; GO:0016740; GO:0008250; GO:0016021
Bottom	chr28	1800001	1900000	g21330.t1	PREDICTED: von Willebrand factor A domain-containing protein 5A-like	
Bottom	chr28	1800001	1900000	g21331.t1	PREDICTED: epimerase family protein SDR39U1 isoform X1	GO:0003824
Bottom	chr28	1800001	1900000	g21332.t1	olfactory receptor 6A2-like	GO:0007186; GO:0050911; GO:0004930; GO:0004984; GO:0005886; GO:0016021
Bottom	chr29	5000001	5100000	g20985.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chr29	5000001	5100000	g20986.t1	hypothetical protein	

					llap_10261	
Bottom	chr29	5000001	5100000	g20987.t1	caytaxin isoform X2	GO:0005737; GO:0016021
Bottom	chr29	5000001	5100000	g20988.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0003824
Bottom	chrZ	8000001	8100000	g9599.t1	PREDICTED: LOW QUALITY PROTEIN: uncharacterized protein LOC106049434	
Bottom	chrZ	8000001	8100000	g9600.t1	PREDICTED: uncharacterized protein LOC104832910, partial	
Bottom	chrZ	8300001	8400000	g9609.t1	hypothetical protein llap_18262	
Bottom	chrZ	8300001	8400000	g9610.t1	hypothetical protein llap_5710	
Bottom	chrZ	8300001	8400000	g9611.t1	dtw domain-containing protein 2	GO:0003824
Bottom	chrZ	8300001	8400000	g9612.t1	pecanex-like protein hypothetical protein	GO:0044237; GO:0016772; GO:0043232
Bottom	chrZ	10200001	10300000	g9660.t1	PAS domain-containing protein	GO:0007165; GO:0016020; GO:0016021
Bottom	chrZ	10200001	10300000	g9661.t1	uncharacterized protein LOC106017933	GO:0046872
Bottom	chrZ	11000001	11100000	g9676.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chrZ	11000001	11100000	g9677.t1	Glutamate receptor 3	
Bottom	chrZ	11000001	11100000	g9678.t1	hypothetical protein llap_7601	
Bottom	chrZ	11000001	11100000	g9679.t1	hypothetical protein llap_16894	
Bottom	chrZ	11000001	11100000	g9680.t1	PREDICTED: uncharacterized protein	GO:0015074; GO:0003676

					LOC109279670	
Bottom	chrZ	14700001	14800000	g9802.t1	rna-directed dna polymerase from mobile element jockey-hypothetical protein	GO:0006278; GO:0003964
Bottom	chrZ	14700001	14800000	g9803.t1	S-methyl-5'-thioadenosine phosphorylase isoform X2	GO:0006166; GO:0019509; GO:0017061; GO:0005634; GO:0005829
Bottom	chrZ	14700001	14800000	g9804.t1	hypothetical protein C5E45_35180, partial	GO:0006278; GO:0051382; GO:0003964; GO:0019237; GO:0000776; GO:0000779; GO:0005634
Bottom	chrZ	14700001	14800000	g9805.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chrZ	14700001	14800000	g9806.t1	RNA-directed DNA polymerase from mobile element jockey, partial	GO:0006278; GO:0003964
Bottom	chrZ	14800001	14900000	g9807.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964; GO:0016021
Bottom	chrZ	14800001	14900000	g9808.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003723; GO:0003964; GO:0005737
Bottom	chrZ	14800001	14900000	g9809.t1	hypothetical protein C5F51_36540, partial	GO:0006278; GO:0003964
Bottom	chrZ	14800001	14900000	g9810.t1	PREDICTED: cyclin-H	GO:0003824
Bottom	chrZ	19700001	19800000	g9952.t1	Coiled-coil domain-containing protein 112, partial	
Bottom	chrZ	19700001	19800000	g9953.t1	PREDICTED: geranylgeranyl transferase type-1 subunit beta isoform X2	GO:0018344; GO:0004662; GO:0005515; GO:0008270; GO:0005953
Bottom	chrZ	19700001	19800000	g9954.t1	E3 ubiquitin-protein ligase TRIM36 isoform X1	GO:0016567; GO:0051726; GO:0004842; GO:0008270; GO:0005737; GO:0005874
Bottom	chrZ	20900001	21000000	g9975.t1	PREDICTED: RNA-directed DNA polymerase from mobile element jockey-like,	GO:0006278; GO:0003964

					partial	
Bottom	chrZ	20900001	21000000	g9976.t1	-	
Bottom	chrZ	24000001	24100000	g10078.t1	hypothetical protein llap_5901	
Bottom	chrZ	24000001	24100000	g10079.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chrZ	24000001	24100000	g10080.t1	hypothetical protein llap_15995	GO:0006278; GO:0006486; GO:0042908; GO:0046677; GO:0055085; GO:0097502; GO:0003824; GO:0003964; GO:0015297; GO:0016757; GO:0042910; GO:0005886; GO:0016020; GO:0016021
Bottom	chrZ	24000001	24100000	g10081.t1	uncharacterized protein LOC106017933	GO:0046872
Bottom	chrZ	34600001	34700000	g10379.t1	hypothetical protein llap_234	
Bottom	chrZ	34600001	34700000	g10380.t1	hypothetical protein llap_399	GO:0006278; GO:0003964
Bottom	chrZ	34600001	34700000	g10381.t1	rna-directed dna polymerase from mobile element jockey-like	GO:0006278; GO:0003964
Bottom	chrZ	38200001	38300000	g10492.t1	glycerol kinase	GO:0016310; GO:0016301
Bottom	chrZ	38200001	38300000	g10493.t1	recQ-mediated genome instability protein 1	GO:0000724; GO:0002023; GO:0006260; GO:0009749; GO:0035264; GO:0042593; GO:0071139; GO:0000166; GO:0016604; GO:0031422
Bottom	chrZ	38200001	38300000	g10494.t1	PREDICTED: heterogeneous nuclear ribonucleoprotein K isoform X2	GO:0010988; GO:0043066; GO:0045892; GO:0045944; GO:0048025; GO:0048260; GO:1902165; GO:1905599; GO:0003677; GO:0003729; GO:0019904; GO:0042802; GO:0000785; GO:0005654; GO:0010494; GO:0071013
Bottom	chrZ	55500001	55600000	g11087.t1	PREDICTED: AP-3 complex subunit beta-1 isoform X2	GO:0000902; GO:0002224; GO:0002244; GO:0003016; GO:0006622; GO:0006882; GO:0006954; GO:0007040; GO:0007283; GO:0007338; GO:0007596; GO:0016182; GO:0030851; GO:0032438; GO:0034394; GO:0042789; GO:0045944; GO:0048007; GO:0048490; GO:0048872;

						GO:0050790; GO:0051138; GO:0060155; GO:0060425; GO:0090152; GO:0098773; GO:0019903; GO:0030742; GO:0005739; GO:0030123; GO:0030131; GO:0030665; GO:0045202; GO:1904115
Bottom	chrZ	55500001	55600000	g11088.t1	hypothetical protein, partial	GO:0006278; GO:0007166; GO:0007186; GO:0003964; GO:0004930; GO:0016021
Bottom	chrZ	55500001	55600000	g11089.t1	AP-3 complex subunit beta-1	GO:0000902; GO:0002224; GO:0002244; GO:0003016; GO:0006622; GO:0006882; GO:0006954; GO:0007040; GO:0007283; GO:0007338; GO:0007596; GO:0016182; GO:0030851; GO:0032438; GO:0034394; GO:0042789; GO:0045944; GO:0048007; GO:0048490; GO:0048872; GO:0050790; GO:0051138; GO:0060155; GO:0060425; GO:0090152; GO:0098773; GO:0019903; GO:0030742; GO:0005739; GO:0030123; GO:0030131; GO:0030665; GO:0045202; GO:1904115
Bottom	chrZ	55500001	55600000	g11090.t1	hypothetical protein, partial	GO:0006278; GO:0003964
Bottom	chrZ	59600001	59700000	g11192.t1	glycerol kinase	GO:0016310; GO:0016301
Bottom	chrZ	59600001	59700000	g11193.t1	pol- hypothetical protein	GO:0006278; GO:0003964
Bottom	chrZ	69900001	70000000	g11489.t1	PREDICTED: dnaJ homolog subfamily B member 5	GO:0051085; GO:0051082; GO:0051087; GO:0005829
Bottom	chrZ	69900001	70000000	g11490.t1	PREDICTED: protein KIAA1045 homolog, partial	GO:0007214; GO:0008277; GO:0032228; GO:0050966; GO:0005509
Bottom	chrZ	69900001	70000000	g11491.t1	PHD finger protein 24	GO:0007214; GO:0008277; GO:0032228; GO:0050966; GO:0005509
Bottom	chrZ	69900001	70000000	g11492.t1	hypothetical protein DUI87_19235	GO:0006812; GO:0065008; GO:0090304; GO:0005215; GO:0016787; GO:0097159; GO:0140657; GO:1901363; GO:0016020
Bottom	chrZ	69900001	70000000	g11493.t1	uncharacterized protein LOC113840406	

Table S3.5. GO Term enrichment analyses for 194 Autosome genes and 130 Z-linked genes for the 1% highest F_{ST}. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be FDR- adjusted p values.

Chromosome Type	GO Term	Fuction	GO Description	p-value	Genes
Autosome	GO:0016020	cc	membrane	No	g3081.t1,g14224.t1
Autosome	GO:0003964	mf	RNA-directed DNA polymerase activity	Yes	g1221.t1,g1261.t1,g1276.t1,g1277.t1,g2359.t1,g2831.t1,g3082.t1,g3083.t1,g3515.t1,g3517.t1,g5971.t1,g5972.t1,g6140.t1,g6141.t1,g6143.t1,g6350.t1,g6351.t1,g6394.t1,g6396.t1,g6402.t1,g6417.t1,g6418.t1,g17272.t1
Autosome	GO:0006278	bp	RNA-templated DNA biosynthetic process	Yes	g1221.t1,g1261.t1,g1276.t1,g1277.t1,g2359.t1,g2831.t1,g3082.t1,g3083.t1,g3515.t1,g3517.t1,g5971.t1,g5972.t1,g6140.t1,g6141.t1,g6143.t1,g6350.t1,g6351.t1,g6394.t1,g6396.t1,g6402.t1,g6417.t1,g6418.t1,g17272.t1
Autosome	GO:0016301	mf	kinase activity	No	g5971.t1
Autosome	GO:0016310	bp	phosphorylation	No	g5971.t1
Autosome	GO:0015935	cc	small ribosomal subunit	No	
Autosome	GO:0003735	mf	structural constituent of ribosome	No	

Autosome	GO:0006412	bp	translation	No	
Autosome	GO:0005789	cc	endoplasmic reticulum membrane	No	
Autosome	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	No	g3516.t1,g6350.t1,g6405.t1,g6417.t1,g6418.t1,g14797.t1,g14800.t1,g14801.t1,g15372.t1,g15575.t1,g17271.t1,g19840.t1,g22433.t1
Autosome	GO:0003677	mf	DNA binding	No	
Autosome	GO:0005524	mf	ATP binding	No	g14924.t1
Autosome	GO:0102158	mf	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase activity	No	
Autosome	GO:0102343	mf	obsolete 3-hydroxy-arachidoyl-CoA dehydratase activity	No	
Autosome	GO:0102344	mf	obsolete 3-hydroxy-behenoyl-CoA dehydratase activity	No	
Autosome	GO:0102345	mf	obsolete 3-hydroxy-lignoceroyl-CoA dehydratase activity	No	
Autosome	GO:0006633	bp	fatty acid biosynthetic process	No	
Autosome	GO:0032508	bp	DNA duplex unwinding	No	
Autosome	GO:0008076	cc	voltage-gated potassium channel complex	No	
Autosome	GO:0005249	mf	voltage-gated potassium channel activity	No	
Autosome	GO:0034765	bp	regulation of monoatomic ion transmembrane transport	No	g14304.t1
Autosome	GO:0051260	bp	protein homooligomerization	No	
Autosome	GO:0071805	bp	potassium ion transmembrane transport	No	
Autosome	GO:0005576	cc	extracellular region	No	g3516.t1,g6352.t1,g14921.t1
Autosome	GO:0005634	cc	nucleus	No	g14224.t1
Autosome	GO:0005737	cc	cytoplasm	No	g3516.t1,g5972.t1,g6306.t1,g14224.t1
Autosome	GO:0005179	mf	hormone activity	Yes	g6352.t1
Autosome	GO:0051428	mf	peptide hormone receptor binding	No	
Autosome	GO:0002076	bp	osteoblast development	No	

Autosome	GO:0003420	bp	regulation of growth plate cartilage chondrocyte proliferation	Yes	
Autosome	GO:0007189	bp	adenylate cyclase-activating G protein-coupled receptor signaling pathway	No	
Autosome	GO:0030282	bp	bone mineralization	No	
Autosome	GO:1902733	bp	regulation of growth plate cartilage chondrocyte differentiation	Yes	
Autosome	GO:1903042	bp	negative regulation of chondrocyte hypertrophy	Yes	
Autosome	GO:0019934	bp	cGMP-mediated signaling	No	
Autosome	GO:0060282	bp	positive regulation of oocyte development	Yes	
Autosome	GO:0040020	bp	regulation of meiotic nuclear division	No	
Autosome	GO:0009410	bp	response to xenobiotic stimulus	No	
Autosome	GO:0043116	bp	negative regulation of vascular permeability	No	
Autosome	GO:0043951	bp	negative regulation of cAMP-mediated signaling	No	
Autosome	GO:0004115	mf	3',5'-cyclic-AMP phosphodiesterase activity	No	
Autosome	GO:0043117	bp	positive regulation of vascular permeability	Yes	
Autosome	GO:0071560	bp	cellular response to transforming growth factor beta stimulus	No	
Autosome	GO:0001556	bp	oocyte maturation	Yes	
Autosome	GO:0071321	bp	cellular response to cGMP	Yes	
Autosome	GO:0046872	mf	metal ion binding	No	
Autosome	GO:0005829	cc	cytosol	No	g14924.t1,g17274.t1
Autosome	GO:0005173	mf	stem cell factor receptor binding	Yes	
Autosome	GO:0008284	bp	positive regulation of cell population proliferation	No	
Autosome	GO:0048513	bp	animal organ development	No	
Autosome	GO:0007155	bp	cell adhesion	No	
Autosome	GO:0045597	bp	positive regulation of cell differentiation	No	
Autosome	GO:1902533	bp	positive regulation of intracellular signal transduction	Yes	
Autosome	GO:0048468	bp	cell development	No	
Autosome	GO:0005125	mf	cytokine activity	No	g14922.t1
Autosome	GO:0005886	cc	plasma membrane	No	g15575.t1,g17274.t1

Autosome	GO:0002684	bp	positive regulation of immune system process	Yes	
Autosome	GO:0003006	bp	developmental process involved in reproduction	Yes	
Autosome	GO:0048731	bp	system development	No	
Autosome	GO:0043066	bp	negative regulation of apoptotic process	No	g14924.t1
Autosome	GO:0001934	bp	positive regulation of protein phosphorylation	No	
Autosome	GO:0030175	cc	filopodium	No	
Autosome	GO:0030027	cc	lamellipodium	No	
Autosome	GO:0003723	mf	RNA binding	No	
Autosome	GO:0005488	mf	binding	No	
Autosome	GO:0005615	cc	extracellular space	No	g6351.t1,g14922.t1
Autosome	GO:0004867	mf	serine-type endopeptidase inhibitor activity	No	g6351.t1
Autosome	GO:0010951	bp	negative regulation of endopeptidase activity	No	g6351.t1
Autosome	GO:0016740	mf	transferase activity	No	
Autosome	GO:0000166	mf	nucleotide binding	No	
Autosome	GO:0000287	mf	magnesium ion binding	No	
Autosome	GO:0106411	mf	XMP 5'-nucleosidase activity	Yes	
Autosome	GO:0009117	bp	nucleotide metabolic process	No	
Autosome	GO:0016311	bp	dephosphorylation	No	
Autosome	GO:0005783	cc	endoplasmic reticulum	No	
Autosome	GO:0003755	mf	peptidyl-prolyl cis-trans isomerase activity	No	
Autosome	GO:0005509	mf	calcium ion binding	No	g19840.t1
Autosome	GO:0000413	bp	protein peptidyl-prolyl isomerization	No	
Autosome	GO:0006457	bp	protein folding	No	
Autosome	GO:0008380	bp	RNA splicing	Yes	g15370.t1
Autosome	GO:0008083	mf	growth factor activity	Yes	g14921.t1,g14922.t1
Autosome	GO:0007165	bp	signal transduction	No	g14921.t1,g14922.t1 ,g14925.t1
Autosome	GO:0033010	cc	paranodal junction	No	
Autosome	GO:0071205	bp	protein localization to juxtaparanode region of axon	No	

Autosome	GO:0004102	mf	choline O-acetyltransferase activity	Yes	
Autosome	GO:0060416	bp	response to growth hormone	No	
Autosome	GO:0016746	mf	acyltransferase activity	No	
Autosome	GO:0042910	mf	xenobiotic transmembrane transporter activity	No	
Autosome	GO:0006635	bp	fatty acid beta-oxidation	Yes	g15371.t1
Autosome	GO:0006836	bp	neurotransmitter transport	No	
Autosome	GO:0042908	bp	xenobiotic transport	No	
Autosome	GO:0055085	bp	transmembrane transport	No	
Autosome	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	No	
Autosome	GO:0004930	mf	G protein-coupled receptor activity	No	
Autosome	GO:0009881	mf	photoreceptor activity	No	
Autosome	GO:0007186	bp	G protein-coupled receptor signaling pathway	No	
Autosome	GO:0007601	bp	visual perception	No	
Autosome	GO:0007602	bp	phototransduction	No	
Autosome	GO:0008236	mf	serine-type peptidase activity	No	
Autosome	GO:0006508	bp	proteolysis	No	g19840.t1
Autosome	GO:0001649	bp	osteoblast differentiation	No	
Autosome	GO:0031281	bp	positive regulation of cyclase activity	Yes	
Autosome	GO:0034198	bp	cellular response to amino acid starvation	No	
Autosome	GO:0031063	bp	regulation of histone deacetylation	No	
Autosome	GO:0090398	bp	cellular senescence	No	
Autosome	GO:1902595	bp	regulation of DNA replication origin binding	Yes	
Autosome	GO:0038095	bp	Fc-epsilon receptor signaling pathway	Yes	
Autosome	GO:0009411	bp	response to UV	No	
Autosome	GO:0045202	cc	synapse	No	
Autosome	GO:0042752	bp	regulation of circadian rhythm	No	
Autosome	GO:0032091	bp	negative regulation of protein binding	No	
Autosome	GO:0007258	bp	JUN phosphorylation	Yes	
Autosome	GO:0051090	bp	regulation of DNA-binding transcription factor activity	Yes	
Autosome	GO:0005654	cc	nucleoplasm	No	g15371.t1,g17274.t1
Autosome	GO:0097441	cc	basal dendrite	Yes	

Autosome	GO:0051247	bp	positive regulation of protein metabolic process	Yes	
Autosome	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	No	
Autosome	GO:0071276	bp	cellular response to cadmium ion	No	
Autosome	GO:0120283	mf	protein serine/threonine kinase binding	Yes	
Autosome	GO:0030424	cc	axon	No	
Autosome	GO:0106310	mf	protein serine kinase activity	No	
Autosome	GO:0035033	mf	histone deacetylase regulator activity	Yes	
Autosome	GO:0019903	mf	protein phosphatase binding	No	
Autosome	GO:0042826	mf	histone deacetylase binding	No	
Autosome	GO:0090045	bp	positive regulation of deacetylase activity	Yes	
Autosome	GO:1900740	bp	positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway	Yes	
Autosome	GO:0071260	bp	cellular response to mechanical stimulus	No	
Autosome	GO:0071222	bp	cellular response to lipopolysaccharide	No	
Autosome	GO:0010628	bp	positive regulation of gene expression	No	
Autosome	GO:0005739	cc	mitochondrion	No	g15371.t1
Autosome	GO:0004705	mf	JUN kinase activity	Yes	
Autosome	GO:0043065	bp	positive regulation of apoptotic process	No	
Autosome	GO:0031343	bp	positive regulation of cell killing	Yes	
Autosome	GO:0016241	bp	regulation of macroautophagy	Yes	
Autosome	GO:0018105	bp	peptidyl-serine phosphorylation	No	
Autosome	GO:0034614	bp	cellular response to reactive oxygen species	No	
Autosome	GO:0018107	bp	peptidyl-threonine phosphorylation	No	
Autosome	GO:0098978	cc	glutamatergic synapse	No	
Autosome	GO:0005096	mf	GTPase activator activity	No	
Autosome	GO:0050790	bp	regulation of catalytic activity	No	g20088.t1
Autosome	GO:0099175	bp	regulation of postsynapse organization	No	
Autosome	GO:0001972	mf	retinoic acid binding	No	
Autosome	GO:0005506	mf	iron ion binding	No	g17271.t1
Autosome	GO:0008401	mf	retinoic acid 4-hydroxylase activity	Yes	
Autosome	GO:0016709	mf	oxidoreductase activity, acting on paired donors, with	No	

			incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen		
Autosome	GO:0020037	mf	heme binding	No	g17271.t1
Autosome	GO:0062183	mf	all-trans retinoic acid 18-hydroxylase activity	Yes	
Autosome	GO:0006805	bp	xenobiotic metabolic process	No	
Autosome	GO:0034653	bp	retinoic acid catabolic process	Yes	
Autosome	GO:0000978	mf	RNA polymerase II cis-regulatory region sequence-specific DNA binding	No	
Autosome	GO:0003700	mf	DNA-binding transcription factor activity	No	
Autosome	GO:0000122	bp	negative regulation of transcription by RNA polymerase II	No	
Autosome	GO:0000902	bp	cell morphogenesis	No	
Autosome	GO:0002070	bp	epithelial cell maturation	Yes	
Autosome	GO:0007028	bp	cytoplasm organization	Yes	
Autosome	GO:0007431	bp	salivary gland development	Yes	
Autosome	GO:0008340	bp	determination of adult lifespan	Yes	
Autosome	GO:0045927	bp	positive regulation of growth	No	
Autosome	GO:0001518	cc	voltage-gated sodium channel complex	No	
Autosome	GO:0005244	mf	voltage-gated monoatomic ion channel activity	No	
Autosome	GO:0005248	mf	voltage-gated sodium channel activity	No	
Autosome	GO:0035725	bp	sodium ion transmembrane transport	No	
Autosome	GO:0005681	cc	spliceosomal complex	No	
Autosome	GO:0035145	cc	exon-exon junction complex	No	
Autosome	GO:0008270	mf	zinc ion binding	No	
Autosome	GO:0005730	cc	nucleolus	No	
Autosome	GO:0004095	mf	carnitine O-palmitoyltransferase activity	Yes	
Autosome	GO:0001676	bp	long-chain fatty acid metabolic process	No	
Autosome	GO:0009437	bp	carnitine metabolic process	Yes	
Autosome	GO:0120162	bp	positive regulation of cold-induced thermogenesis	No	
Autosome	GO:0015293	mf	symporter activity	No	
Autosome	GO:0015807	bp	L-amino acid transport	No	
Autosome	GO:0098656	bp	monoatomic anion transmembrane transport	Yes	g15575.t1

Autosome	GO:1902475	bp	L-alpha-amino acid transmembrane transport	No	
Autosome	GO:0045162	bp	clustering of voltage-gated sodium channels	Yes	
Autosome	GO:0004497	mf	monooxygenase activity	No	
Autosome	GO:0016705	mf	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	No	
Autosome	GO:0008526	mf	phosphatidylinositol transfer activity	Yes	
Autosome	GO:0035091	mf	phosphatidylinositol binding	No	
Autosome	GO:0015914	bp	phospholipid transport	No	
Autosome	GO:0042981	bp	regulation of apoptotic process	No	
Autosome	GO:0043552	bp	positive regulation of phosphatidylinositol 3-kinase activity	No	
Autosome	GO:0048017	bp	inositol lipid-mediated signaling	Yes	
Autosome	GO:0051897	bp	positive regulation of protein kinase B signaling	No	
Autosome	GO:0070374	bp	positive regulation of ERK1 and ERK2 cascade	No	
Autosome	GO:0120009	bp	intermembrane lipid transfer	No	
Autosome	GO:0005085	mf	guanyl-nucleotide exchange factor activity	No	
Autosome	GO:0007264	bp	small GTPase mediated signal transduction	No	
Autosome	GO:0005802	cc	trans-Golgi network	No	
Autosome	GO:0031410	cc	cytoplasmic vesicle	No	
Autosome	GO:0006886	bp	intracellular protein transport	No	
Autosome	GO:0099041	bp	vesicle tethering to Golgi	Yes	
Autosome	GO:0004252	mf	serine-type endopeptidase activity	No	
Z	GO:0005634	cc	nucleus	No	g10940.t1,g11004.t1
Z	GO:0005739	cc	mitochondrion	No	g11025.t1
Z	GO:0003729	mf	mRNA binding	No	
Z	GO:0004146	mf	dihydrofolate reductase activity	Yes	
Z	GO:0050661	mf	NADP binding	No	
Z	GO:0006545	bp	glycine biosynthetic process	Yes	
Z	GO:0006729	bp	tetrahydrobiopterin biosynthetic process	No	
Z	GO:0031427	bp	response to methotrexate	Yes	
Z	GO:0035999	bp	tetrahydrofolate interconversion	No	

Z	GO:0046654	bp	tetrahydrofolate biosynthetic process	Yes	
Z	GO:0051000	bp	positive regulation of nitric-oxide synthase activity	No	
Z	GO:2000121	bp	regulation of removal of superoxide radicals	No	
Z	GO:0016020	cc	membrane	No	g9782.t1,g10930.t1
Z	GO:0016301	mf	kinase activity	No	
Z	GO:0008360	bp	regulation of cell shape	No	
Z	GO:0016310	bp	phosphorylation	No	
Z	GO:0009986	cc	cell surface	No	
Z	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	No	g9658.t1,g9664.t1,g9784.t1,g10502.t1,g10951.t1,g11111.t1
Z	GO:0022857	mf	transmembrane transporter activity	No	g10502.t1
Z	GO:0043029	bp	T cell homeostasis	No	
Z	GO:0045580	bp	regulation of T cell differentiation	Yes	
Z	GO:0048538	bp	thymus development	No	
Z	GO:0055085	bp	transmembrane transport	No	g10502.t1
Z	GO:0070233	bp	negative regulation of T cell apoptotic process	Yes	
Z	GO:0070430	bp	positive regulation of nucleotide-binding oligomerization domain containing 1 signaling pathway	Yes	
Z	GO:0005576	cc	extracellular region	No	g10502.t1
Z	GO:0005654	cc	nucleoplasm	No	
Z	GO:0005829	cc	cytosol	No	
Z	GO:0042803	mf	protein homodimerization activity	No	
Z	GO:0047134	mf	protein-disulfide reductase (NAD(P)) activity	Yes	
Z	GO:0000122	bp	negative regulation of transcription by RNA polymerase II	No	
Z	GO:0009314	bp	response to radiation	No	
Z	GO:0033138	bp	positive regulation of peptidyl-serine phosphorylation	No	
Z	GO:0043388	bp	positive regulation of DNA binding	No	
Z	GO:0045454	bp	cell redox homeostasis	No	
Z	GO:0046826	bp	negative regulation of protein export from nucleus	Yes	
Z	GO:0071731	bp	response to nitric oxide	Yes	

Z	GO:1903206	bp	negative regulation of hydrogen peroxide-induced cell death	Yes	
Z	GO:2000170	bp	positive regulation of peptidyl-cysteine S-nitrosylation	Yes	
Z	GO:0045211	cc	postsynaptic membrane	No	
Z	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	No	
Z	GO:0004714	mf	transmembrane receptor protein tyrosine kinase activity	No	
Z	GO:0005524	mf	ATP binding	No	g9663.t1,g9782.t1,g10502.t1
Z	GO:0008582	bp	regulation of synaptic assembly at neuromuscular junction	Yes	
Z	GO:0018108	bp	peptidyl-tyrosine phosphorylation	No	g10010.t1
Z	GO:0030154	bp	cell differentiation	No	
Z	GO:0004383	mf	guanylate cyclase activity	No	
Z	GO:0004672	mf	protein kinase activity	No	g10502.t1
Z	GO:0006182	bp	cGMP biosynthetic process	No	
Z	GO:0006468	bp	protein phosphorylation	No	g10502.t1
Z	GO:0007601	bp	visual perception	No	
Z	GO:0035556	bp	intracellular signal transduction	No	
Z	GO:0031981	cc	nuclear lumen	No	
Z	GO:0005488	mf	binding	No	
Z	GO:0016787	mf	hydrolase activity	No	g11027.t1
Z	GO:0006259	bp	DNA metabolic process	No	
Z	GO:0006974	bp	DNA damage response	No	
Z	GO:0030544	mf	Hsp70 protein binding	No	
Z	GO:0046872	mf	metal ion binding	No	g10940.t1,g11004.t1
Z	GO:0050750	mf	low-density lipoprotein particle receptor binding	No	
Z	GO:0051082	mf	unfolded protein binding	No	
Z	GO:0006457	bp	protein folding	No	
Z	GO:0007283	bp	spermatogenesis	No	
Z	GO:0009408	bp	response to heat	No	
Z	GO:0030317	bp	flagellated sperm motility	No	

Z	GO:0030521	bp	androgen receptor signaling pathway	No	
Z	GO:0000381	bp	regulation of alternative mRNA splicing, via spliceosome	No	
Z	GO:0005737	cc	cytoplasm	No	
Z	GO:0016607	cc	nuclear speck	No	
Z	GO:0071005	cc	U2-type precatalytic spliceosome	No	
Z	GO:0032580	cc	Golgi cisterna membrane	No	
Z	GO:0016757	mf	glycosyltransferase activity	No	
Z	GO:0005975	bp	carbohydrate metabolic process	No	
Z	GO:0006486	bp	protein glycosylation	No	
Z	GO:0003824	mf	catalytic activity	No	g10510.t1
Z	GO:0004713	mf	protein tyrosine kinase activity	No	
Z	GO:0005154	mf	epidermal growth factor receptor binding	No	
Z	GO:0008283	bp	cell population proliferation	No	
Z	GO:0030335	bp	positive regulation of cell migration	No	
Z	GO:0031532	bp	actin cytoskeleton reorganization	No	
Z	GO:0042058	bp	regulation of epidermal growth factor receptor signaling pathway	No	
Z	GO:0051092	bp	positive regulation of NF-kappaB transcription factor activity	No	
Z	GO:0003964	mf	RNA-directed DNA polymerase activity	Yes	g10298.t1,g10600.t1,g10601.t1,g10744.t1,g10745.t1,g10892.t1,g10894.t1,g10927.t1,g10928.t1,g10929.t1,g10930.t1,g10999.t1,g11000.t1,g11254.t1,g11599.t1,g11600.t1,g11603.t1

Z	GO:0006278	bp	RNA-templated DNA biosynthetic process	Yes	g10298.t1,g10600.t1,g10601.t1,g10744.t1,g10745.t1,g10892.t1,g10894.t1,g10927.t1,g10928.t1,g10929.t1,g10930.t1,g10999.t1,g11000.t1,g11254.t1,g11599.t1,g11600.t1,g11603.t1
Z	GO:0003676	mf	nucleic acid binding	No	
Z	GO:0015074	bp	DNA integration	No	
Z	GO:0004930	mf	G protein-coupled receptor activity	No	
Z	GO:0004984	mf	olfactory receptor activity	No	
Z	GO:0005540	mf	hyaluronic acid binding	No	
Z	GO:0007155	bp	cell adhesion	No	
Z	GO:0007186	bp	G protein-coupled receptor signaling pathway	No	
Z	GO:0050911	bp	detection of chemical stimulus involved in sensory perception of smell	No	
Z	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	No	
Z	GO:0005227	mf	calcium activated cation channel activity	No	
Z	GO:0005262	mf	calcium channel activity	No	
Z	GO:0016048	bp	detection of temperature stimulus	No	
Z	GO:0050951	bp	sensory perception of temperature stimulus	No	
Z	GO:0051262	bp	protein tetramerization	No	
Z	GO:0070588	bp	calcium ion transmembrane transport	No	
Z	GO:0016740	mf	transferase activity	No	
Z	GO:0007049	bp	cell cycle	No	
Z	GO:0016567	bp	protein ubiquitination	No	
Z	GO:0009897	cc	external side of plasma membrane	No	
Z	GO:0006955	bp	immune response	No	
Z	GO:0007166	bp	cell surface receptor signaling pathway	No	

Z	GO:0031295	bp	T cell costimulation	No	
Z	GO:0042102	bp	positive regulation of T cell proliferation	No	
Z	GO:0042130	bp	negative regulation of T cell proliferation	No	
Z	GO:0071222	bp	cellular response to lipopolysaccharide	No	
Z	GO:0000977	mf	RNA polymerase II transcription regulatory region sequence-specific DNA binding	No	
Z	GO:0001228	mf	DNA-binding transcription activator activity, RNA polymerase II-specific	No	
Z	GO:0042802	mf	identical protein binding	No	
Z	GO:0014807	bp	regulation of somitogenesis	No	
Z	GO:0045944	bp	positive regulation of transcription by RNA polymerase II	No	
Z	GO:0048706	bp	embryonic skeletal system development	No	
Z	GO:2000287	bp	positive regulation of myotome development	Yes	
Z	GO:0050684	bp	regulation of mRNA processing	No	
Z	GO:0003677	mf	DNA binding	No	g11598.t1
Z	GO:0003700	mf	DNA-binding transcription factor activity	No	g11598.t1
Z	GO:0006355	bp	regulation of DNA-templated transcription	No	g11598.t1
Z	GO:0010824	bp	regulation of centrosome duplication	No	
Z	GO:0005905	cc	clathrin-coated pit	No	
Z	GO:0072583	bp	clathrin-dependent endocytosis	No	
Z	GO:0005794	cc	Golgi apparatus	No	
Z	GO:0005385	mf	zinc ion transmembrane transporter activity	No	
Z	GO:0006882	bp	intracellular zinc ion homeostasis	No	
Z	GO:0071577	bp	zinc ion transmembrane transport	No	
Z	GO:0004222	mf	metalloendopeptidase activity	No	
Z	GO:0008270	mf	zinc ion binding	No	
Z	GO:0006508	bp	proteolysis	No	

Table S3.6. GO Term enrichment analyses for 360 Autosome genes and106 Z-linked genes under for the 1% lowest F_{ST}. Significant differences were tested by Fisher's exact test and shown by p values, and they are corrected to be FDR- adjusted p values.

Chromosome Type	GO Term	Fuction	GO Description	p-value	Genes

Autosomes	GO:0003964	mf	RNA-directed DNA polymerase activity	Yes	g301.t1,g306.t1,g710.t1,g711.t1,g713.t1,g714.t1,g715.t1,g1184.t1,g1186.t1,g1492.t1,g2088.t1,g2090.t1,g2091.t1,g2092.t1,g2722.t1,g2840.t1,g2842.t1,g2902.t1,g2903.t1,g2904.t1,g2957.t1,g2958.t1,g2961.t1,g2962.t1,g3018.t1,g3090.t1,g3459.t1,g3460.t1,g3813.t1,g3816.t1,g3952.t1,g4106.t1,g4107.t1,g4110.t1,g4111.t1,g4117.t1,g4727.t1,g4729.t1,g4731.t1,g4763.t1,g4764.t1,g4765.t1,g4766.t1,g5057.t1,g5064.t1,g5066.t1,g5068.t1,g5071.t1,g5252.t1,g5253.t1,g5254.t1,g5258.t1,g5319.t1,g5786.t1,g5800.t1,g6147.t1,g6148.t1,g8275.t1,g8277.t1,g8279.t1,g8280.t1,g8282.t1,g8477.t1,g8479.t1,g8653.t1,g8852.t1,g8853.t1,g8854.t1,g8858.t1,g9181.t1,g9182.t1,g9222.t1,g9223.t1,g9224.t1,g11868.t1,g11918.t1,g12022.t1,g12024.t1,g12027.t1,g12028.t1,g12548.t1,g12567.t1,g13066.t1,g13195.t1,g13196.t1,g13198.t1,g13724.t1,g15569.t1,g15614.t1,g15616.t1,g15623.t1,g15625.t1,g15626.t1,g15636.t1,g15640.t1,g15644.t1,g15645.t1,g15657.t1,g15801.t1,g15911.t1,g15928.t1,g17708.t1,g17709.t1,g17275.t1,g17290.t1,g17291.t1,g17543.t1,g18944.t1,g19344.t1,g20037.t1,g20040.t1,g20044.t1,g20052.t1,g20061.t1,g20698.t1,g20719.t1,g20720.t1,g21875.t1,g21876.t1,g21247.t1,g21248.t1,g21250.t1,g21251.t1,g21252.t1,g21253.t1,g21254.t1,g21255.t1,g21256.t1,g21258.t1,g21270.t1,g21324.t1

Autosomes	GO:0006278	bp	RNA-templated DNA biosynthetic process	Yes	g301.t1,g306.t1,g710.t1,g711.t1,g713.t1,g714.t1,g715.t1,g1184.t1,g1186.t1,g1492.t1,g2088.t1,g2090.t1,g2091.t1,g2092.t1,g2722.t1,g2840.t1,g2842.t1,g2902.t1,g2903.t1,g2904.t1,g2957.t1,g2958.t1,g2961.t1,g2962.t1,g3018.t1,g3090.t1,g3459.t1,g3460.t1,g3813.t1,g3816.t1,g3952.t1,g4106.t1,g4107.t1,g4110.t1,g4111.t1,g4117.t1,g4727.t1,g4729.t1,g4731.t1,g4763.t1,g4764.t1,g4765.t1,g4766.t1,g5057.t1,g5064.t1,g5066.t1,g5068.t1,g5071.t1,g5252.t1,g5253.t1,g5254.t1,g5258.t1,g5319.t1,g5786.t1,g5800.t1,g6147.t1,g6148.t1,g8275.t1,g8277.t1,g8279.t1,g8280.t1,g8282.t1,g8477.t1,g8479.t1,g8653.t1,g8852.t1,g8853.t1,g8854.t1,g8858.t1,g9181.t1,g9182.t1,g9222.t1,g9223.t1,g9224.t1,g11868.t1,g11918.t1,g12022.t1,g12024.t1,g12027.t1,g12028.t1,g12548.t1,g12567.t1,g13066.t1,g13195.t1,g13196.t1,g13198.t1,g13724.t1,g15569.t1,g15614.t1,g15616.t1,g15623.t1,g15625.t1,g15626.t1,g15636.t1,g15640.t1,g15644.t1,g15645.t1,g15657.t1,g15801.t1,g15911.t1,g15928.t1,g17708.t1,g17709.t1,g17275.t1,g17290.t1,g17291.t1,g17543.t1,g18944.t1,g19344.t1,g20037.t1,g20040.t1,g20044.t1,g20052.t1,g20061.t1,g20698.t1,g20719.t1,g20720.t1,g21875.t1,g21876.t1,g21247.t1,g21248.t1,g21250.t1,g21251.t1,g21252.t1,g21253.t1,g21254.t1,g21255.t1,g21256.t1,g21258.t1,g21270.t1,g21324.t1

Autosomes	GO:0003824	mf	catalytic activity	Yes	g302.t1,g304.t1,g712.t1,g1495.t1,g2959.t1,g3456.t1,g3811.t1,g3820.t1,g4116.t1,g4306.t1,g5060.t1,g5320.t1,g6880.t1,g8131.t1,g8276.t1,g13723.t1,g14841.t1,g15615.t1,g15629.t1,g15632.t1,g17285.t1,g16431.t1,g19345.t1,g20043.t1,g21259.t1,g21331.t1,g20988.t1
Autosomes	GO:0003677	mf	DNA binding	No	g701.t1,g1184.t1,g2841.t1,g20704.t1
Autosomes	GO:0004386	mf	helicase activity	No	g701.t1,g19335.t1
Autosomes	GO:0005524	mf	ATP binding	No	g701.t1,g2703.t1,g2723.t1,g2842.t1,g8273.t1,g14339.t1,g19335.t1,g19337.t1,g20043.t1,g20047.t1,g20050.t1,g20054.t1,g20055.t1,g20059.t1,g20060.t1,g20062.t1,g20063.t1,g20064.t1,g20065.t1,g20704.t1
Autosomes	GO:0008170	mf	N-methyltransferase activity	No	g701.t1
Autosomes	GO:0016787	mf	hydrolase activity	No	g701.t1,g2953.t1,g2979.t1,g6881.t1
Autosomes	GO:0006306	bp	DNA methylation	No	g701.t1
Autosomes	GO:0007218	bp	neuropeptide signaling pathway	No	g709.t1
Autosomes	GO:0005615	cc	extracellular space	No	g711.t1,g713.t1,g2658.t1,g2726.t1,g2727.t1,g2903.t1,g4110.t1,g20701.t1,g21875.t1
Autosomes	GO:0004867	mf	serine-type endopeptidase inhibitor activity	No	g711.t1,g713.t1,g2903.t1,g4110.t1,g6148.t1,g8275.t1,g21875.t1
Autosomes	GO:0010951	bp	negative regulation of endopeptidase activity	No	g711.t1,g713.t1,g2658.t1,g2903.t1,g4110.t1,g6148.t1,g8275.t1,g21875.t1
Autosomes	GO:0004523	mf	RNA-DNA hybrid ribonuclease activity	No	g1184.t1,g3019.t1,g3458.t1,g12027.t1,g21250.t1
Autosomes	GO:0008233	mf	peptidase activity	No	g1184.t1
Autosomes	GO:0008270	mf	zinc ion binding	No	g1184.t1,g2661.t1,g2841.t1,g3014.t1,g5319.t1,g15914.t1,g17157.t1,g16499.t1,g20703.t1,g22115.t1,g21250.t1
Autosomes	GO:0015074	bp	DNA integration	No	g1184.t1,g3458.t1,g12027.t1,g21250.t1
Autosomes	GO:0090502	bp	RNA phosphodiester bond hydrolysis,	No	g1184.t1,g3019.t1,g3458.t1,g12027.t1,g16

			endonucleolytic		498.t1,g21250.t1
Autosomes	GO:0110165	cc	cellular anatomical entity	No	g1497.t1,g2945.t1,g3453.t1,g3934.t1,g17286.t1
Autosomes	GO:0031514	cc	motile cilium	No	g2024.t1
Autosomes	GO:0016020	cc	membrane	No	g2340.t1,g8282.t1,g8849.t1,g12017.t1,g13196.t1,g18082.t1,g18084.t1,g17279.t1,g17280.t1,g17289.t1,g22778.t1,g21554.t1
Autosomes	GO:0033644	cc	host cell membrane	No	g2340.t1
Autosomes	GO:0019058	bp	viral life cycle	No	g2340.t1
Autosomes	GO:0051701	bp	biological process involved in interaction with host	No	g2340.t1
Autosomes	GO:0005730	cc	nucleolus	No	g2658.t1,g17710.t1,g18082.t1,g18084.t1,g20722.t1
Autosomes	GO:0005829	cc	cytosol	Yes	g2658.t1,g2660.t1,g2946.t1,g17710.t1,g18082.t1,g18084.t1,g18129.t1,g19337.t1
Autosomes	GO:0002020	mf	protease binding	No	g2658.t1
Autosomes	GO:0004869	mf	cysteine-type endopeptidase inhibitor activity	No	g2658.t1
Autosomes	GO:0005758	cc	mitochondrial intermembrane space	No	g2659.t1
Autosomes	GO:0005654	cc	nucleoplasm	No	g2660.t1,g3014.t1,g17710.t1,g20722.t1,g19209.t1,g22394.t1
Autosomes	GO:0014069	cc	postsynaptic density	No	g2660.t1
Autosomes	GO:0030425	cc	dendrite	No	g2660.t1,g15650.t1
Autosomes	GO:0042564	cc	NLS-dependent protein nuclear import complex	No	g2660.t1
Autosomes	GO:0098978	cc	glutamatergic synapse	No	g2660.t1
Autosomes	GO:0008139	mf	nuclear localization sequence binding	No	g2660.t1
Autosomes	GO:0061608	mf	nuclear import signal receptor activity	No	g2660.t1,g19829.t1
Autosomes	GO:0006607	bp	NLS-bearing protein import into nucleus	No	g2660.t1
Autosomes	GO:0014841	bp	skeletal muscle satellite cell proliferation	No	g2660.t1
Autosomes	GO:0014901	bp	satellite cell activation involved in skeletal muscle regeneration	No	g2660.t1
Autosomes	GO:0042981	bp	regulation of apoptotic process	No	g2660.t1
Autosomes	GO:0060828	bp	regulation of canonical Wnt signaling pathway	No	g2660.t1

Autosomes	GO:0099527	bp	postsynapse to nucleus signaling pathway	No	g2660.t1
Autosomes	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	No	g2661.t1,g2702.t1,g2841.t1,g2904.t1,g2963.t1,g3017.t1,g3021.t1,g3022.t1,g3454.t1,g3459.t1,g3952.t1,g4727.t1,g4728.t1,g5064.t1,g5258.t1,g5319.t1,g6880.t1,g8273.t1,g9183.t1,g12030.t1,g13195.t1,g15534.t1,g15636.t1,g15659.t1,g15915.t1,g16349.t1,g17711.t1,g17712.t1,g17277.t1,g17281.t1,g17282.t1,g17541.t1,g18201.t1,g18202.t1,g18203.t1,g18204.t1,g18205.t1,g19336.t1,g19339.t1,g19340.t1,g19346.t1,g22774.t1,g22779.t1,g22780.t1,g22781.t1,g22782.t1,g22783.t1,g22784.t1,g20086.t1,g20699.t1,g20721.t1,g20723.t1,g19210.t1,g22115.t1,g22392.t1,g21559.t1,g21246.t1,g21252.t1,g21270.t1,g21273.t1,g21329.t1,g21332.t1,g20987.t1
Autosomes	GO:0015280	mf	ligand-gated sodium channel activity	No	g2661.t1
Autosomes	GO:0061630	mf	ubiquitin protein ligase activity	No	g2661.t1,g2841.t1,g19209.t1
Autosomes	GO:0002376	bp	immune system process	No	g2661.t1
Autosomes	GO:0016567	bp	protein ubiquitination	No	g2661.t1,g2841.t1
Autosomes	GO:0035725	bp	sodium ion transmembrane transport	No	g2661.t1,g21246.t1
Autosomes	GO:0005886	cc	plasma membrane	No	g2702.t1,g12030.t1,g15915.t1,g15927.t1,g17710.t1,g17714.t1,g17281.t1,g17541.t1,g18201.t1,g18202.t1,g18203.t1,g18204.t1,g18205.t1,g20725.t1,g19209.t1,g21332.t1
Autosomes	GO:0004930	mf	G protein-coupled receptor activity	No	g2702.t1,g2904.t1,g21332.t1
Autosomes	GO:0004984	mf	olfactory receptor activity	No	g2702.t1,g21332.t1
Autosomes	GO:0007186	bp	G protein-coupled receptor signaling pathway	No	g2702.t1,g2904.t1,g17277.t1,g21332.t1
Autosomes	GO:0050911	bp	detection of chemical stimulus involved in sensory perception of smell	No	g2702.t1,g21332.t1
Autosomes	GO:0005887	cc	obsolete proteoglycan integral to plasma membrane	No	g2703.t1,g2720.t1,g2722.t1,g2723.t1

Autosomes	GO:0004712	mf	protein serine/threonine/tyrosine kinase activity	No	g2703.t1,g2723.t1
Autosomes	GO:0005003	mf	ephrin receptor activity	No	g2703.t1,g2723.t1
Autosomes	GO:0018108	bp	peptidyl-tyrosine phosphorylation	No	g2703.t1,g2723.t1
Autosomes	GO:0048013	bp	ephrin receptor signaling pathway	No	g2703.t1,g2723.t1
Autosomes	GO:0005262	mf	calcium channel activity	No	g2720.t1,g2722.t1
Autosomes	GO:0098703	bp	calcium ion import across plasma membrane	No	g2720.t1,g2722.t1
Autosomes	GO:0009986	cc	cell surface	No	g2723.t1
Autosomes	GO:0001806	bp	type IV hypersensitivity	No	g2723.t1
Autosomes	GO:0032092	bp	positive regulation of protein binding	No	g2723.t1
Autosomes	GO:0050798	bp	activated T cell proliferation	No	g2723.t1
Autosomes	GO:2000525	bp	positive regulation of T cell costimulation	No	g2723.t1
Autosomes	GO:0004252	mf	serine-type endopeptidase activity	No	g2726.t1,g2727.t1,g16428.t1,g20725.t1
Autosomes	GO:0046872	mf	metal ion binding	No	g2726.t1,g2727.t1,g3019.t1,g3458.t1,g8853.t1,g12027.t1,g13906.t1,g17279.t1,g17280.t1,g16496.t1,g22786.t1,g21272.t1
Autosomes	GO:0006508	bp	proteolysis	No	g2726.t1,g2727.t1,g2841.t1,g3019.t1,g12027.t1,g15914.t1,g16428.t1,g22115.t1,g22391.t1
Autosomes	GO:0007586	bp	digestion	No	g2726.t1,g2727.t1
Autosomes	GO:0016772	mf	transferase activity, transferring phosphorus-containing groups	No	g2730.t1,g3453.t1,g21266.t1
Autosomes	GO:0044237	bp	cellular metabolic process	No	g2730.t1,g21266.t1
Autosomes	GO:0003676	mf	nucleic acid binding	No	g2840.t1,g3019.t1,g3458.t1,g5319.t1,g8853.t1,g16498.t1,g19335.t1,g22393.t1,g21250.t1
Autosomes	GO:0000118	cc	histone deacetylase complex	No	g2841.t1
Autosomes	GO:0005737	cc	cytoplasm	Yes	g2841.t1,g2859.t1,g2952.t1,g2980.t1,g3935.t1,g4766.t1,g15804.t1,g16499.t1,g18130.t1,g18494.t1,g19829.t1,g20987.t1
Autosomes	GO:0003714	mf	transcription corepressor activity	No	g2841.t1
Autosomes	GO:0004198	mf	calcium-dependent cysteine-type	No	g2841.t1

			endopeptidase activity		
Autosomes	GO:0006357	bp	regulation of transcription by RNA polymerase II	No	g2841.t1,g13906.t1,g16499.t1
Autosomes	GO:0006397	bp	mRNA processing	No	g2841.t1
Autosomes	GO:0016575	bp	histone deacetylation	No	g2841.t1
Autosomes	GO:0045892	bp	negative regulation of DNA-templated transcription	No	g2841.t1
Autosomes	GO:0055085	bp	transmembrane transport	No	g2841.t1,g8273.t1,g20086.t1
Autosomes	GO:0003723	mf	RNA binding	No	g2842.t1,g4766.t1,g12027.t1,g14401.t1,g21251.t1,g21252.t1
Autosomes	GO:0003724	mf	RNA helicase activity	No	g2842.t1
Autosomes	GO:0005085	mf	guanyl-nucleotide exchange factor activity	No	g2859.t1,g7370.t1,g18679.t1
Autosomes	GO:0035556	bp	intracellular signal transduction	No	g2859.t1
Autosomes	GO:0050790	bp	regulation of catalytic activity	No	g2859.t1,g2980.t1,g7370.t1,g12475.t1,g18131.t1,g18679.t1,g20085.t1,g21329.t1
Autosomes	GO:0007166	bp	cell surface receptor signaling pathway	No	g2904.t1
Autosomes	GO:0005739	cc	mitochondrion	No	g2946.t1,g19336.t1
Autosomes	GO:0008146	mf	sulfotransferase activity	No	g2946.t1
Autosomes	GO:0042802	mf	identical protein binding	No	g2946.t1,g16429.t1
Autosomes	GO:0006790	bp	sulfur compound metabolic process	No	g2946.t1
Autosomes	GO:0005925	cc	focal adhesion	No	g2952.t1
Autosomes	GO:0015629	cc	actin cytoskeleton	No	g2952.t1
Autosomes	GO:0003779	mf	actin binding	No	g2952.t1
Autosomes	GO:0007163	bp	establishment or maintenance of cell polarity	No	g2952.t1
Autosomes	GO:0030031	bp	cell projection assembly	No	g2952.t1
Autosomes	GO:0031532	bp	actin cytoskeleton reorganization	No	g2952.t1
Autosomes	GO:0034446	bp	substrate adhesion-dependent cell spreading	No	g2952.t1
Autosomes	GO:0005488	mf	binding	No	g2953.t1,g4107.t1

Autosomes	GO:0044260	bp	obsolete cellular macromolecule metabolic process	No	g2953.t1,g3019.t1
Autosomes	GO:0090304	bp	nucleic acid metabolic process	No	g2953.t1
Autosomes	GO:0009987	bp	cellular process	No	g2959.t1,g3934.t1,g4762.t1,g8271.t1,g12017.t1,g15641.t1,g17285.t1
Autosomes	GO:0044238	bp	primary metabolic process	No	g2959.t1
Autosomes	GO:0071704	bp	organic substance metabolic process	No	g2959.t1
Autosomes	GO:0005509	mf	calcium ion binding	No	g2963.t1,g3017.t1,g3021.t1,g3454.t1,g12030.t1,g13916.t1,g17281.t1,g17541.t1,g18201.t1,g18202.t1,g18203.t1,g18204.t1,g18205.t1
Autosomes	GO:0005576	cc	extracellular region	No	g2964.t1,g4295.t1,g4731.t1,g8273.t1,g21274.t1
Autosomes	GO:0009374	mf	biotin binding	No	g2964.t1,g4731.t1
Autosomes	GO:0016301	mf	kinase activity	No	g2964.t1,g4731.t1,g5863.t1,g12018.t1,g13916.t1,g21247.t1
Autosomes	GO:0016310	bp	phosphorylation	No	g2964.t1,g4731.t1,g5863.t1,g12018.t1,g13916.t1,g21247.t1
Autosomes	GO:0005096	mf	GTPase activator activity	No	g2980.t1,g7370.t1,g12475.t1,g20085.t1
Autosomes	GO:0007264	bp	small GTPase mediated signal transduction	No	g2980.t1
Autosomes	GO:2001136	bp	negative regulation of endocytic recycling	No	g2980.t1
Autosomes	GO:1902894	bp	negative regulation of miRNA transcription	No	g3014.t1
Autosomes	GO:0032099	bp	negative regulation of appetite	No	g3014.t1
Autosomes	GO:2001171	bp	positive regulation of ATP biosynthetic process	No	g3014.t1
Autosomes	GO:0000978	mf	RNA polymerase II cis-regulatory region sequence-specific DNA binding	No	g3014.t1,g13906.t1,g17710.t1
Autosomes	GO:0008544	bp	epidermis development	No	g3014.t1
Autosomes	GO:0045820	bp	negative regulation of glycolytic process	No	g3014.t1
Autosomes	GO:1900016	bp	negative regulation of cytokine production involved in inflammatory response	No	g3014.t1
Autosomes	GO:0008289	mf	lipid binding	No	g3014.t1
Autosomes	GO:0001228	mf	DNA-binding transcription activator activity, RNA	No	g3014.t1

			polymerase II-specific		
Autosomes	GO:0001227	mf	DNA-binding transcription repressor activity, RNA polymerase II-specific	No	g3014.t1,g17710.t1
Autosomes	GO:0045944	bp	positive regulation of transcription by RNA polymerase II	No	g3014.t1
Autosomes	GO:0042752	bp	regulation of circadian rhythm	No	g3014.t1
Autosomes	GO:0043401	bp	steroid hormone mediated signaling pathway	No	g3014.t1
Autosomes	GO:1903427	bp	negative regulation of reactive oxygen species biosynthetic process	No	g3014.t1
Autosomes	GO:0045722	bp	positive regulation of gluconeogenesis	No	g3014.t1
Autosomes	GO:1903944	bp	negative regulation of hepatocyte apoptotic process	No	g3014.t1
Autosomes	GO:0000122	bp	negative regulation of transcription by RNA polymerase II	No	g3014.t1,g17710.t1
Autosomes	GO:0032922	bp	circadian regulation of gene expression	No	g3014.t1
Autosomes	GO:0042060	bp	wound healing	No	g3014.t1
Autosomes	GO:0046321	bp	positive regulation of fatty acid oxidation	No	g3014.t1
Autosomes	GO:0004879	mf	nuclear receptor activity	No	g3014.t1,g16499.t1
Autosomes	GO:2000112	bp	obsolete regulation of cellular macromolecule biosynthetic process	No	g3014.t1,g16499.t1
Autosomes	GO:0003707	mf	nuclear steroid receptor activity	No	g3014.t1
Autosomes	GO:0045070	bp	positive regulation of viral genome replication	No	g3014.t1
Autosomes	GO:0030522	bp	intracellular receptor signaling pathway	No	g3014.t1,g16499.t1
Autosomes	GO:0010891	bp	negative regulation of sequestering of triglyceride	No	g3014.t1
Autosomes	GO:2000272	bp	negative regulation of signaling receptor activity	No	g3014.t1
Autosomes	GO:0050728	bp	negative regulation of inflammatory response	No	g3014.t1
Autosomes	GO:1903038	bp	negative regulation of leukocyte cell-cell adhesion	No	g3014.t1
Autosomes	GO:0009267	bp	cellular response to starvation	No	g3014.t1,g17714.t1
Autosomes	GO:0046889	bp	positive regulation of lipid biosynthetic process	No	g3014.t1
Autosomes	GO:0031624	mf	ubiquitin conjugating enzyme binding	No	g3014.t1
Autosomes	GO:0061629	mf	RNA polymerase II-specific DNA-binding	No	g3014.t1

			transcription factor binding		
Autosomes	GO:0010745	bp	negative regulation of macrophage derived foam cell differentiation	No	g3014.t1
Autosomes	GO:0010887	bp	negative regulation of cholesterol storage	No	g3014.t1
Autosomes	GO:0030512	bp	negative regulation of transforming growth factor beta receptor signaling pathway	No	g3014.t1
Autosomes	GO:0051898	bp	negative regulation of protein kinase B signaling	No	g3014.t1
Autosomes	GO:0004190	mf	aspartic-type endopeptidase activity	No	g3019.t1,g12027.t1,g22391.t1
Autosomes	GO:0045202	cc	synapse	No	g3287.t1
Autosomes	GO:0030354	mf	melanin-concentrating hormone activity	No	g3287.t1
Autosomes	GO:0007165	bp	signal transduction	No	g3287.t1,g7370.t1,g12475.t1,g13196.t1,g20701.t1
Autosomes	GO:0007268	bp	chemical synaptic transmission	No	g3287.t1
Autosomes	GO:0005905	cc	clathrin-coated pit	No	g3454.t1
Autosomes	GO:0034361	cc	very-low-density lipoprotein particle	No	g3454.t1
Autosomes	GO:0030229	mf	very-low-density lipoprotein particle receptor activity	No	g3454.t1
Autosomes	GO:0006869	bp	lipid transport	No	g3454.t1,g18082.t1,g18084.t1
Autosomes	GO:0006897	bp	endocytosis	No	g3454.t1
Autosomes	GO:0008203	bp	cholesterol metabolic process	No	g3454.t1
Autosomes	GO:0022414	bp	reproductive process	No	g3934.t1
Autosomes	GO:0007283	bp	spermatogenesis	No	g3935.t1,g21272.t1
Autosomes	GO:0030154	bp	cell differentiation	No	g3935.t1
Autosomes	GO:0031047	bp	RNA-mediated gene silencing	No	g3935.t1
Autosomes	GO:0051321	bp	meiotic cell cycle	No	g3935.t1
Autosomes	GO:0004518	mf	nuclease activity	No	g4107.t1
Autosomes	GO:0030215	mf	semaphorin receptor binding	No	g4295.t1
Autosomes	GO:0016032	bp	viral process	Yes	g6148.t1,g21251.t1,g21252.t1

Autosomes	GO:0030275	mf	LRR domain binding	No	g6324.t1
Autosomes	GO:0006355	bp	regulation of DNA-templated transcription	No	g6324.t1,g17713.t1,g22772.t1,g22786.t1
Autosomes	GO:0004350	mf	glutamate-5-semialdehyde dehydrogenase activity	Yes	g7368.t1,g7369.t1
Autosomes	GO:0055129	bp	L-proline biosynthetic process	No	g7368.t1,g7369.t1
Autosomes	GO:0050661	mf	NADP binding	No	g7369.t1
Autosomes	GO:0016192	bp	vesicle-mediated transport	No	g7370.t1,g21273.t1
Autosomes	GO:0004672	mf	protein kinase activity	No	g8273.t1,g14339.t1
Autosomes	GO:0005540	mf	hyaluronic acid binding	No	g8273.t1,g18130.t1
Autosomes	GO:0022857	mf	transmembrane transporter activity	No	g8273.t1,g20086.t1
Autosomes	GO:0006468	bp	protein phosphorylation	No	g8273.t1,g14339.t1
Autosomes	GO:0007155	bp	cell adhesion	No	g8273.t1
Autosomes	GO:0005794	cc	Golgi apparatus	No	g8282.t1,g21273.t1
Autosomes	GO:0005515	mf	protein binding	No	g8282.t1
Autosomes	GO:0005871	cc	kinesin complex	No	g8478.t1
Autosomes	GO:0019894	mf	kinesin binding	No	g8478.t1
Autosomes	GO:0004519	mf	endonuclease activity	No	g8853.t1
Autosomes	GO:0090305	bp	nucleic acid phosphodiester bond hydrolysis	No	g8853.t1,g18945.t1
Autosomes	GO:0016829	mf	lyase activity	No	g12017.t1,g16498.t1
Autosomes	GO:0043167	mf	ion binding	No	g12017.t1
Autosomes	GO:0003887	mf	DNA-directed DNA polymerase activity	No	g12027.t1,g21251.t1,g21252.t1
Autosomes	GO:0006310	bp	DNA recombination	No	g12027.t1
Autosomes	GO:0007156	bp	homophilic cell adhesion via plasma membrane adhesion molecules	No	g12030.t1,g17281.t1,g17541.t1,g18201.t1, g18202.t1,g18203.t1,g18204.t1,g18205.t1
Autosomes	GO:0004888	mf	transmembrane signaling receptor activity	No	g13196.t1
Autosomes	GO:0005634	cc	nucleus	Yes	g13906.t1,g15804.t1,g15913.t1,g16429.t1, g16499.t1,g18131.t1,g18945.t1,g20724.t1
Autosomes	GO:0005667	cc	transcription regulator complex	No	g13906.t1

Autosomes	GO:0045666	bp	positive regulation of neuron differentiation	No	g15648.t1,g15650.t1
Autosomes	GO:0045930	bp	negative regulation of mitotic cell cycle	No	g15648.t1,g15650.t1
Autosomes	GO:0005783	cc	endoplasmic reticulum	No	g15650.t1
Autosomes	GO:0043025	cc	neuronal cell body	No	g15650.t1
Autosomes	GO:0071300	bp	cellular response to retinoic acid	No	g15650.t1
Autosomes	GO:0005516	mf	calmodulin binding	No	g15804.t1,g20060.t1
Autosomes	GO:0007051	bp	spindle organization	No	g15804.t1
Autosomes	GO:0005846	cc	nuclear cap binding complex	No	g15913.t1
Autosomes	GO:0000339	mf	RNA cap binding	No	g15913.t1
Autosomes	GO:0045292	bp	mRNA cis splicing, via spliceosome	No	g15913.t1
Autosomes	GO:0004181	mf	metallocarboxypeptidase activity	No	g15914.t1
Autosomes	GO:0001596	mf	angiotensin type I receptor activity	No	g15915.t1
Autosomes	GO:0004945	mf	angiotensin type II receptor activity	No	g15915.t1
Autosomes	GO:0031711	mf	bradykinin receptor binding	No	g15915.t1
Autosomes	GO:0046982	mf	protein heterodimerization activity	No	g15915.t1,g20722.t1
Autosomes	GO:0001822	bp	kidney development	No	g15915.t1
Autosomes	GO:0006954	bp	inflammatory response	No	g15915.t1
Autosomes	GO:0007266	bp	Rho protein signal transduction	No	g15915.t1
Autosomes	GO:0010873	bp	obsolete positive regulation of cholesterol esterification	No	g15915.t1
Autosomes	GO:0019229	bp	regulation of vasoconstriction	No	g15915.t1
Autosomes	GO:0019722	bp	calcium-mediated signaling	No	g15915.t1
Autosomes	GO:0032430	bp	positive regulation of phospholipase A2 activity	No	g15915.t1
Autosomes	GO:0046718	bp	viral entry into host cell	No	g15915.t1
Autosomes	GO:0051482	bp	positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G protein-coupled signaling pathway	No	g15915.t1
Autosomes	GO:0060326	bp	cell chemotaxis	No	g15915.t1,g20701.t1

Autosomes	GO:0086097	bp	phospholipase C-activating angiotensin-activated signaling pathway	No	g15915.t1
Autosomes	GO:1903589	bp	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	No	g15915.t1
Autosomes	GO:0017128	mf	phospholipid scramblase activity	No	g15927.t1
Autosomes	GO:0017121	bp	plasma membrane phospholipid scrambling	No	g15927.t1
Autosomes	GO:0030246	mf	carbohydrate binding	No	g16349.t1
Autosomes	GO:0090599	mf	alpha-glucosidase activity	No	g16349.t1
Autosomes	GO:0005975	bp	carbohydrate metabolic process	No	g16349.t1
Autosomes	GO:0008327	mf	methyl-CpG binding	No	g17710.t1
Autosomes	GO:0009617	bp	response to bacterium	No	g17711.t1,g17712.t1
Autosomes	GO:0005637	cc	nuclear inner membrane	No	g17713.t1
Autosomes	GO:0005890	cc	sodium:potassium-exchanging ATPase complex	No	g17713.t1
Autosomes	GO:0006813	bp	potassium ion transport	No	g17713.t1
Autosomes	GO:0006814	bp	sodium ion transport	No	g17713.t1
Autosomes	GO:0005765	cc	lysosomal membrane	No	g17714.t1
Autosomes	GO:0031902	cc	late endosome membrane	No	g17714.t1,g20724.t1
Autosomes	GO:0097637	cc	obsolete integral component of autophagosome membrane	No	g17714.t1
Autosomes	GO:0061740	bp	protein targeting to lysosome involved in chaperone-mediated autophagy	No	g17714.t1
Autosomes	GO:0097352	bp	autophagosome maturation	No	g17714.t1
Autosomes	GO:1905146	bp	lysosomal protein catabolic process	No	g17714.t1
Autosomes	GO:0000938	cc	GARP complex	No	g18082.t1,g18084.t1
Autosomes	GO:1990745	cc	EARP complex	No	g18082.t1,g18084.t1
Autosomes	GO:0006914	bp	autophagy	No	g18082.t1,g18084.t1
Autosomes	GO:0007030	bp	Golgi organization	No	g18082.t1,g18084.t1
Autosomes	GO:0007041	bp	lysosomal transport	No	g18082.t1,g18084.t1
Autosomes	GO:0015031	bp	protein transport	No	g18082.t1,g18084.t1,g20724.t1,g21273.t1
Autosomes	GO:0032456	bp	endocytic recycling	No	g18082.t1,g18084.t1
Autosomes	GO:0042147	bp	retrograde transport, endosome to Golgi	No	g18082.t1,g18084.t1

Autosomes	GO:0048193	bp	Golgi vesicle transport	No	g18082.t1,g18084.t1
Autosomes	GO:0048854	bp	brain morphogenesis	No	g18082.t1,g18084.t1
Autosomes	GO:0004843	mf	cysteine-type deubiquitinase activity	No	g17157.t1,g18678.t1
Autosomes	GO:0006511	bp	ubiquitin-dependent protein catabolic process	No	g17157.t1,g18678.t1
Autosomes	GO:0016579	bp	protein deubiquitination	No	g17157.t1,g18678.t1
Autosomes	GO:0001750	cc	photoreceptor outer segment	No	g17277.t1
Autosomes	GO:0009968	bp	negative regulation of signal transduction	No	g17277.t1
Autosomes	GO:0050908	bp	detection of light stimulus involved in visual perception	No	g17277.t1
Autosomes	GO:0098609	bp	cell-cell adhesion	No	g17279.t1,g17280.t1
Autosomes	GO:0016757	mf	glycosyltransferase activity	No	g17282.t1,g21247.t1
Autosomes	GO:0050896	bp	response to stimulus	No	g17285.t1
Autosomes	GO:0005911	cc	cell-cell junction	No	g17541.t1
Autosomes	GO:0008013	mf	beta-catenin binding	No	g17541.t1
Autosomes	GO:0046983	mf	protein dimerization activity	No	g17541.t1
Autosomes	GO:0030199	bp	collagen fibril organization	No	g17541.t1
Autosomes	GO:0001726	cc	ruffle	No	g16429.t1
Autosomes	GO:0031901	cc	early endosome membrane	No	g16429.t1
Autosomes	GO:0045335	cc	phagocytic vesicle	No	g16429.t1
Autosomes	GO:0007049	bp	cell cycle	No	g16429.t1
Autosomes	GO:0016740	mf	transferase activity	No	g16496.t1,g18131.t1,g18493.t1,g22393.t1, g21329.t1
Autosomes	GO:0000214	cc	tRNA-intron endonuclease complex	No	g16498.t1
Autosomes	GO:0000213	mf	tRNA-intron endonuclease activity	No	g16498.t1
Autosomes	GO:0006388	bp	tRNA splicing, via endonucleolytic cleavage and ligation	No	g16498.t1
Autosomes	GO:0043565	mf	sequence-specific DNA binding	No	g16499.t1
Autosomes	GO:0048511	bp	rhythmic process	No	g16499.t1
Autosomes	GO:0045171	cc	intercellular bridge	No	g18129.t1

Autosomes	GO:0072686	cc	mitotic spindle	No	g18129.t1
Autosomes	GO:0005856	cc	cytoskeleton	No	g18130.t1
Autosomes	GO:0048269	cc	methionine adenosyltransferase complex	No	g18131.t1
Autosomes	GO:0019899	mf	enzyme binding	No	g18131.t1
Autosomes	GO:0048270	mf	methionine adenosyltransferase regulator activity	No	g18131.t1
Autosomes	GO:0006556	bp	S-adenosylmethionine biosynthetic process	No	g18131.t1
Autosomes	GO:0006730	bp	one-carbon metabolic process	No	g18131.t1
Autosomes	GO:0106388	mf	18S rRNA aminocarboxypropyltransferase activity	No	g18494.t1
Autosomes	GO:1904047	mf	S-adenosyl-L-methionine binding	No	g18494.t1
Autosomes	GO:0000455	bp	enzyme-directed rRNA pseudouridine synthesis	No	g18494.t1
Autosomes	GO:0032012	bp	regulation of ARF protein signal transduction	No	g18679.t1
Autosomes	GO:0003697	mf	single-stranded DNA binding	No	g18945.t1
Autosomes	GO:0004520	mf	DNA endonuclease activity	No	g18945.t1
Autosomes	GO:0006281	bp	DNA repair	No	g18945.t1
Autosomes	GO:0030729	mf	acetoacetate-CoA ligase activity	No	g19337.t1
Autosomes	GO:0006631	bp	fatty acid metabolic process	No	g19337.t1
Autosomes	GO:0032024	bp	positive regulation of insulin secretion	No	g19337.t1
Autosomes	GO:0065007	bp	biological regulation	No	g20084.t1
Autosomes	GO:0051056	bp	regulation of small GTPase mediated signal transduction	No	g20085.t1
Autosomes	GO:0008643	bp	carbohydrate transport	No	g20086.t1
Autosomes	GO:0006606	bp	protein import into nucleus	No	g19829.t1
Autosomes	GO:0016459	cc	myosin complex	Yes	g20043.t1,g20047.t1,g20050.t1,g20054.t1, g20055.t1,g20059.t1,g20060.t1,g20062.t1, g20063.t1,g20064.t1,g20065.t1
Autosomes	GO:0003774	mf	cytoskeletal motor activity	Yes	g20043.t1,g20047.t1,g20050.t1,g20054.t1, g20055.t1,g20059.t1,g20060.t1,g20062.t1, g20063.t1,g20064.t1,g20065.t1
Autosomes	GO:0051015	mf	actin filament binding	Yes	g20043.t1,g20047.t1,g20050.t1,g20054.t1, g20055.t1,g20059.t1,g20060.t1,g20062.t1, g20063.t1,g20064.t1,g20065.t1

Autosomes	GO:0030016	cc	myofibril	No	g20047.t1,g20060.t1
Autosomes	GO:0032982	cc	myosin filament	No	g20060.t1
Autosomes	GO:0003841	mf	1-acylglycerol-3-phosphate O-acyltransferase activity	No	g20699.t1
Autosomes	GO:0008654	bp	phospholipid biosynthetic process	No	g20699.t1
Autosomes	GO:0004601	mf	peroxidase activity	No	g20700.t1
Autosomes	GO:0020037	mf	heme binding	No	g20700.t1
Autosomes	GO:0006979	bp	response to oxidative stress	No	g20700.t1
Autosomes	GO:0098869	bp	cellular oxidant detoxification	No	g20700.t1
Autosomes	GO:0008009	mf	chemokine activity	No	g20701.t1
Autosomes	GO:0006955	bp	immune response	No	g20701.t1
Autosomes	GO:0032508	bp	DNA duplex unwinding	No	g20704.t1
Autosomes	GO:0006836	bp	neurotransmitter transport	No	g20721.t1
Autosomes	GO:0061025	bp	membrane fusion	No	g20721.t1,g21273.t1
Autosomes	GO:0005484	mf	SNAP receptor activity	No	g20721.t1,g21273.t1
Autosomes	GO:0000149	mf	SNARE binding	No	g20721.t1
Autosomes	GO:0017157	bp	regulation of exocytosis	No	g20721.t1
Autosomes	GO:0006886	bp	intracellular protein transport	No	g20721.t1
Autosomes	GO:0048471	cc	perinuclear region of cytoplasm	No	g20722.t1
Autosomes	GO:0016435	mf	rRNA (guanine) methyltransferase activity	Yes	g20722.t1,g20724.t1
Autosomes	GO:0070476	bp	rRNA (guanine-N7)-methylation	Yes	g20722.t1,g20724.t1
Autosomes	GO:2000234	bp	positive regulation of rRNA processing	No	g20722.t1
Autosomes	GO:0000813	cc	ESCRT I complex	No	g20724.t1
Autosomes	GO:0046907	bp	intracellular transport	No	g20724.t1
Autosomes	GO:0031638	bp	zymogen activation	No	g20725.t1
Autosomes	GO:0005769	cc	early endosome	No	g19209.t1
Autosomes	GO:0005938	cc	cell cortex	No	g19209.t1
Autosomes	GO:0032991	cc	protein-containing complex	No	g19209.t1
Autosomes	GO:0016874	mf	ligase activity	No	g19209.t1
Autosomes	GO:0043021	mf	ribonucleoprotein complex binding	No	g19209.t1
Autosomes	GO:0044389	mf	ubiquitin-like protein ligase binding	No	g19209.t1
Autosomes	GO:0045236	mf	CXCR chemokine receptor binding	No	g19209.t1

Autosomes	GO:1990763	mf	arrestin family protein binding	No	g19209.t1
Autosomes	GO:0002669	bp	positive regulation of T cell anergy	No	g19209.t1
Autosomes	GO:0032088	bp	negative regulation of NF-kappaB transcription factor activity	No	g19209.t1
Autosomes	GO:0035519	bp	protein K29-linked ubiquitination	No	g19209.t1
Autosomes	GO:0043066	bp	negative regulation of apoptotic process	No	g19209.t1,g21329.t1
Autosomes	GO:0043161	bp	proteasome-mediated ubiquitin-dependent protein catabolic process	No	g19209.t1
Autosomes	GO:0045732	bp	positive regulation of protein catabolic process	No	g19209.t1
Autosomes	GO:0046329	bp	negative regulation of JNK cascade	No	g19209.t1
Autosomes	GO:0046642	bp	negative regulation of alpha-beta T cell proliferation	No	g19209.t1
Autosomes	GO:0050687	bp	negative regulation of defense response to virus	No	g19209.t1
Autosomes	GO:0051865	bp	protein autoubiquitination	No	g19209.t1
Autosomes	GO:0070534	bp	protein K63-linked ubiquitination	No	g19209.t1
Autosomes	GO:0070936	bp	protein K48-linked ubiquitination	No	g19209.t1
Autosomes	GO:0090085	bp	regulation of protein deubiquitination	No	g19209.t1
Autosomes	GO:2000646	bp	positive regulation of receptor catabolic process	No	g19209.t1
Autosomes	GO:0097250	bp	mitochondrial respirasome assembly	No	g19210.t1
Autosomes	GO:0031012	cc	extracellular matrix	No	g22115.t1
Autosomes	GO:0004222	mf	metalloendopeptidase activity	No	g22115.t1
Autosomes	GO:0000003	bp	reproduction	No	g22115.t1
Autosomes	GO:0071013	cc	catalytic step 2 spliceosome	No	g22394.t1
Autosomes	GO:0003729	mf	mRNA binding	No	g22394.t1
Autosomes	GO:0005882	cc	intermediate filament	Yes	g21551.t1,g21552.t1,g21553.t1,g21555.t1, g21556.t1,g21557.t1,g21257.t1,g21263.t1, g21268.t1
Autosomes	GO:0005200	mf	structural constituent of cytoskeleton	Yes	g21551.t1,g21552.t1,g21553.t1,g21555.t1, g21556.t1,g21557.t1,g21257.t1,g21263.t1, g21268.t1
Autosomes	GO:0007010	bp	cytoskeleton organization	Yes	g21551.t1,g21552.t1,g21553.t1,g21555.t1, g21556.t1,g21557.t1,g21257.t1,g21263.t1, g21268.t1

Autosomes	GO:0005272	mf	sodium channel activity	No	g21246.t1
Autosomes	GO:0042605	mf	peptide antigen binding	No	g21267.t1
Autosomes	GO:0043154	bp	negative regulation of cysteine-type endopeptidase activity involved in apoptotic process	No	g21270.t1
Autosomes	GO:0043027	mf	cysteine-type endopeptidase inhibitor activity involved in apoptotic process	No	g21270.t1
Autosomes	GO:0005102	mf	signaling receptor binding	No	g21274.t1
Autosomes	GO:0016055	bp	Wnt signaling pathway	No	g21274.t1
Autosomes	GO:0008250	cc	oligosaccharyltransferase complex	No	g21329.t1
Autosomes	GO:0031647	bp	regulation of protein stability	No	g21329.t1
Autosomes	GO:0006487	bp	protein N-linked glycosylation	No	g21329.t1
Autosomes	GO:0008047	mf	enzyme activator activity	No	g21329.t1
Z	GO:0003824	mf	catalytic activity	No	g9611.t1,g9810.t1
Z	GO:0043232	cc	intracellular non-membrane-bounded organelle	Yes	g9612.t1
Z	GO:0016772	mf	transferase activity, transferring phosphorus-containing groups	No	g9612.t1
Z	GO:0044237	bp	cellular metabolic process	No	g9612.t1
Z	GO:0003964	mf	RNA-directed DNA polymerase activity	Yes	g9676.t1,g9802.t1,g9805.t1,g9806.t1,g9807.t1,g9808.t1,g9809.t1,g10079.t1,g10380.t1,g10381.t1,g11088.t1,g11090.t1
Z	GO:0006278	bp	RNA-templated DNA biosynthetic process	Yes	g9676.t1,g9802.t1,g9805.t1,g9806.t1,g9807.t1,g9808.t1,g9809.t1,g10079.t1,g10380.t1,g10381.t1,g11088.t1,g11090.t1
Z	GO:0003676	mf	nucleic acid binding	No	g9680.t1
Z	GO:0015074	bp	DNA integration	No	g9680.t1
Z	GO:0005634	cc	nucleus	No	g9803.t1
Z	GO:0005829	cc	cytosol	No	g9803.t1,g11489.t1
Z	GO:0017061	mf	S-methyl-5-thioadenosine phosphorylase activity	Yes	g9803.t1
Z	GO:0006166	bp	purine ribonucleoside salvage	Yes	g9803.t1
Z	GO:0019509	bp	L-methionine salvage from methylthioadenosine	Yes	g9803.t1
Z	GO:0016021	bp	CDP-diacylglycerol biosynthetic process	Yes	g9807.t1,g11088.t1

Z	GO:0005737	cc	cytoplasm	No	g9808.t1,g9954.t1
Z	GO:0003723	mf	RNA binding	No	g9808.t1
Z	GO:0005953	cc	CAAX-protein geranylgeranyltransferase complex	Yes	g9953.t1
Z	GO:0004662	mf	CAAX-protein geranylgeranyltransferase activity	Yes	g9953.t1
Z	GO:0005515	mf	protein binding	No	g9953.t1
Z	GO:0008270	mf	zinc ion binding	No	g9953.t1,g9954.t1
Z	GO:0018344	bp	protein geranylgeranylation	Yes	g9953.t1
Z	GO:0005874	cc	microtubule	No	g9954.t1
Z	GO:0004842	mf	ubiquitin-protein transferase activity	No	g9954.t1
Z	GO:0016567	bp	protein ubiquitination	No	g9954.t1
Z	GO:0051726	bp	regulation of cell cycle	No	g9954.t1
Z	GO:0016301	mf	kinase activity	No	g10492.t1,g11192.t1
Z	GO:0016310	bp	phosphorylation	No	g10492.t1,g11192.t1
Z	GO:0016604	cc	nuclear body	No	g10493.t1
Z	GO:0031422	cc	RecQ family helicase-topoisomerase III complex	Yes	g10493.t1
Z	GO:0000166	mf	nucleotide binding	No	g10493.t1
Z	GO:0000724	bp	double-strand break repair via homologous recombination	No	g10493.t1
Z	GO:0002023	bp	reduction of food intake in response to dietary excess	Yes	g10493.t1
Z	GO:0006260	bp	DNA replication	No	g10493.t1
Z	GO:0009749	bp	response to glucose	Yes	g10493.t1
Z	GO:0035264	bp	multicellular organism growth	No	g10493.t1
Z	GO:0042593	bp	glucose homeostasis	No	g10493.t1
Z	GO:0071139	bp	resolution of recombination intermediates	Yes	g10493.t1
Z	GO:0019904	mf	protein domain specific binding	No	g10494.t1
Z	GO:0042802	mf	identical protein binding	No	g10494.t1
Z	GO:0010494	cc	cytoplasmic stress granule	No	g10494.t1
Z	GO:1905599	bp	positive regulation of low-density lipoprotein receptor activity	Yes	g10494.t1
Z	GO:1902165	bp	regulation of intrinsic apoptotic signaling pathway in response to DNA damage by p53	Yes	g10494.t1

			class mediator		
Z	GO:0045944	bp	positive regulation of transcription by RNA polymerase II	No	g10494.t1,g11087.t1,g11089.t1
Z	GO:0000785	cc	chromatin	No	g10494.t1
Z	GO:0005654	cc	nucleoplasm	No	g10494.t1
Z	GO:0003677	mf	DNA binding	No	g10494.t1
Z	GO:0071013	cc	catalytic step 2 spliceosome	No	g10494.t1
Z	GO:0045892	bp	negative regulation of DNA-templated transcription	No	g10494.t1
Z	GO:0010988	bp	regulation of low-density lipoprotein particle clearance	Yes	g10494.t1
Z	GO:0048025	bp	negative regulation of mRNA splicing, via spliceosome	Yes	g10494.t1
Z	GO:0043066	bp	negative regulation of apoptotic process	No	g10494.t1
Z	GO:0003729	mf	mRNA binding	No	g10494.t1
Z	GO:0048260	bp	positive regulation of receptor-mediated endocytosis	Yes	g10494.t1
Z	GO:0005739	cc	mitochondrion	No	g11087.t1,g11089.t1
Z	GO:0030123	cc	AP-3 adaptor complex	Yes	g11087.t1,g11089.t1
Z	GO:0030131	cc	clathrin adaptor complex	Yes	g11087.t1,g11089.t1
Z	GO:0030665	cc	clathrin-coated vesicle membrane	Yes	g11087.t1,g11089.t1
Z	GO:0045202	cc	synapse	No	g11087.t1,g11089.t1
Z	GO:1904115	cc	axon cytoplasm	Yes	g11087.t1,g11089.t1
Z	GO:0019903	mf	protein phosphatase binding	Yes	g11087.t1,g11089.t1
Z	GO:0030742	mf	GTP-dependent protein binding	Yes	g11087.t1,g11089.t1
Z	GO:0000902	bp	cell morphogenesis	Yes	g11087.t1,g11089.t1
Z	GO:0002224	bp	toll-like receptor signaling pathway	Yes	g11087.t1,g11089.t1
Z	GO:0002244	bp	hematopoietic progenitor cell differentiation	Yes	g11087.t1,g11089.t1
Z	GO:0003016	bp	respiratory system process	Yes	g11087.t1,g11089.t1
Z	GO:0006622	bp	protein targeting to lysosome	Yes	g11087.t1,g11089.t1
Z	GO:0006882	bp	intracellular zinc ion homeostasis	Yes	g11087.t1,g11089.t1

Z	GO:0006954	bp	inflammatory response	Yes	g11087.t1,g11089.t1
Z	GO:0007040	bp	lysosome organization	Yes	g11087.t1,g11089.t1
Z	GO:0007283	bp	spermatogenesis	Yes	g11087.t1,g11089.t1
Z	GO:0007338	bp	single fertilization	Yes	g11087.t1,g11089.t1
Z	GO:0007596	bp	blood coagulation	Yes	g11087.t1,g11089.t1
Z	GO:0016182	bp	synaptic vesicle budding from endosome	Yes	g11087.t1,g11089.t1
Z	GO:0030851	bp	granulocyte differentiation	Yes	g11087.t1,g11089.t1
Z	GO:0032438	bp	melanosome organization	Yes	g11087.t1,g11089.t1
Z	GO:0034394	bp	protein localization to cell surface	Yes	g11087.t1,g11089.t1
Z	GO:0042789	bp	mRNA transcription by RNA polymerase II	Yes	g11087.t1,g11089.t1
Z	GO:0048007	bp	antigen processing and presentation, exogenous lipid antigen via MHC class Ib	Yes	g11087.t1,g11089.t1
Z	GO:0048490	bp	anterograde synaptic vesicle transport	Yes	g11087.t1,g11089.t1
Z	GO:0048872	bp	homeostasis of number of cells	Yes	g11087.t1,g11089.t1
Z	GO:0050790	bp	regulation of catalytic activity	No	g11087.t1,g11089.t1
Z	GO:0051138	bp	positive regulation of NK T cell differentiation	Yes	g11087.t1,g11089.t1
Z	GO:0060155	bp	platelet dense granule organization	Yes	g11087.t1,g11089.t1
Z	GO:0060425	bp	lung morphogenesis	Yes	g11087.t1,g11089.t1
Z	GO:0090152	bp	establishment of protein localization to mitochondrial membrane involved in mitochondrial fission	Yes	g11087.t1,g11089.t1
Z	GO:0098773	bp	skin epidermis development	Yes	g11087.t1,g11089.t1
Z	GO:0004930	mf	G protein-coupled receptor activity	No	g11088.t1
Z	GO:0007166	bp	cell surface receptor signaling pathway	No	g11088.t1
Z	GO:0007186	bp	G protein-coupled receptor signaling pathway	No	g11088.t1
Z	GO:0051082	mf	unfolded protein binding	No	g11489.t1
Z	GO:0051087	mf	protein-folding chaperone binding	No	g11489.t1
Z	GO:0051085	bp	chaperone cofactor-dependent protein refolding	Yes	g11489.t1
Z	GO:0005509	mf	calcium ion binding	No	g11490.t1,g11491.t1
Z	GO:0007214	bp	gamma-aminobutyric acid signaling pathway	Yes	g11490.t1,g11491.t1
Z	GO:0008277	bp	regulation of G protein-coupled receptor signaling pathway	Yes	g11490.t1,g11491.t1

Z	GO:0032228	bp	regulation of synaptic transmission, GABAergic	Yes	g11490.t1,g11491.t1
Z	GO:0050966	bp	detection of mechanical stimulus involved in sensory perception of pain	Yes	g11490.t1,g11491.t1
Z	GO:0016020	cc	membrane	No	g11492.t1
Z	GO:0005215	mf	transporter activity	Yes	g11492.t1
Z	GO:0016787	mf	hydrolase activity	No	g11492.t1
Z	GO:0097159	mf	organic cyclic compound binding	No	g11492.t1
Z	GO:0140657	mf	ATP-dependent activity	Yes	g11492.t1
Z	GO:1901363	mf	heterocyclic compound binding	No	g11492.t1
Z	GO:0006812	bp	monoatomic cation transport	Yes	g11492.t1
Z	GO:0065008	bp	regulation of biological quality	No	g11492.t1
Z	GO:0090304	bp	nucleic acid metabolic process	No	g11492.t1

Table S4.1. Sample information.

Sample Size	Category	Ecoregion	State/ Province	Latitude	Longitude
1	Park Mallard	Hawaii	Hawaii	22.03616	159.77894
1	Hawaiian Duck	Hawaii	Hawaii	22.03616	159.77894
1	Hybrid 1	Hawaii	Hawaii	22.03616	159.77894
1	Hybrid 2	Hawaii	Hawaii	22.03616	159.77894
1	Hybrid 3	Hawaii	Hawaii	22.03616	159.77894
1	New Zeaand Mallard	New Zealand	Canterbury	-44.096992	171.4691
1	New Zealand Grey Duck	New Zealand	West Coast	-42.3122	171.6027
1	Hybrid 1	New Zealand	Tasman	-41.795741	172.3309
1	Hybrid 2	New Zealand	Tasman	-41.411411	173.01607
1	Hybrid 3	New Zealand	West Coast	-42.796218	170.91844
1	Hybrid 4	New Zealand	West Coast	-43.8208	171.6985
1	Wild Mallard	North America	New Mexico	32.953968	-107.29599
1	Game-Farm Mallard	North America	New Jersey	39.626111	-75.486944
1	Hybrid 1	North America	New Jersey	39.8472	-75.0005
1	Hybrid 2	North America	Massachusetts	42.375	-71.375
1	Hybrid 3	North America	Connecticut	41.83671	-72.55994

Table S4.2. Genomic sequencing information.

		Autosomes + Z-sex chromosome			
Ecoregion	Category	No. reads	Covered bases (Mbp)	Alignment (%)	Mean coverage (x)
Hawaii	Hawaiian duck	22332019	34.02	99.1	84.6
	Hybrid 1	4809866	35.52	99.00	18.2
	Hybrid 2	5787065	35.54	99.03	21.2
	Hybrid 3	4789362	35.53	99.02	17.9
	Park Mallard	5482921	35.55	99.10	20.6
New Zealand	New Zealand Grey Duck	4873445	34.00	99.07	20.4
	Hybrid 1	5258468	35.54	99.03	20.9
	Hybrid 2	5379820	35.54	99.09	22.1
	Hybrid 3	5189987	35.55	99.10	21.4
	Hybrid 4	5338833	34.74	99.11	22.6
	New Zealand Mallard	5435162	35.55	99.10	21.2
North America	Wild Mallard	20777591	34.07	99.31	67.4
	Hybrid 1	6960977	35.57	99.15	26.7
	Hybrid 2	6876137	35.57	99.15	26.9
	Hybrid 3	5673930	35.56	99.12	21.9
	Game Farm Mallard	7825476	35.51	99.04	33.0

SUPPLEMENTAL FIGURES

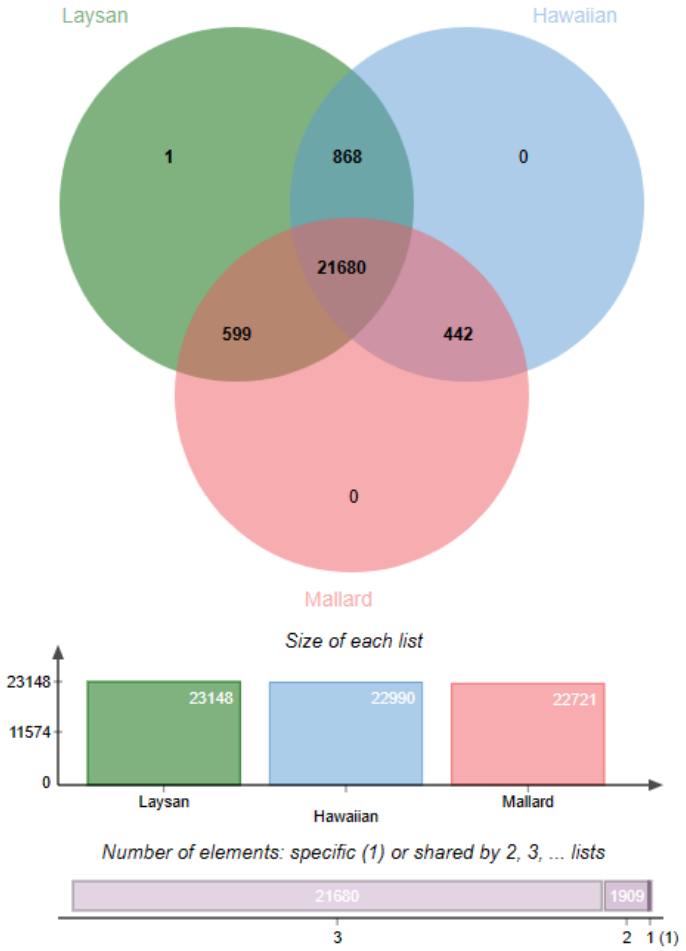

Figure S1.1: (top) OrthVenn2 Venn diagram highlighting the overlap of the protein clusters of each of the duck species. Singletons (i.e. proteins that did not form a cluster and were not counted within the figure) numbered 1,446 for Laysan, 655 for Hawaiian, and 578 for mallard. (middle) Protein cluster counts among the different genomes. The size of the list of proteins in each cluster was balanced suggesting that assembly did not have many duplicated regions. (bottom) Majority of the protein clusters overlapped among all the different taxa (21,680). For the protein clusters between two lists, an increased amount between the Laysan duck and Hawaiian were found (868 of the 1909, or 45%). Followed by the overlap between the Laysan duck and the mallard (599) and the lowest overlap was between the Hawaiian duck and the mallard (442).

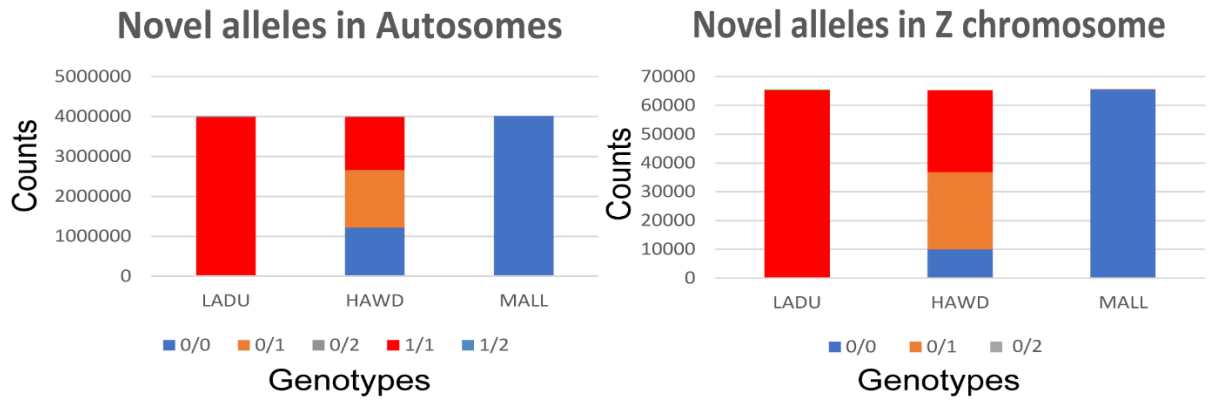

Figure S1.2: Novel Alleles count in Autosomes (left) and the Z-chromosome. (right).

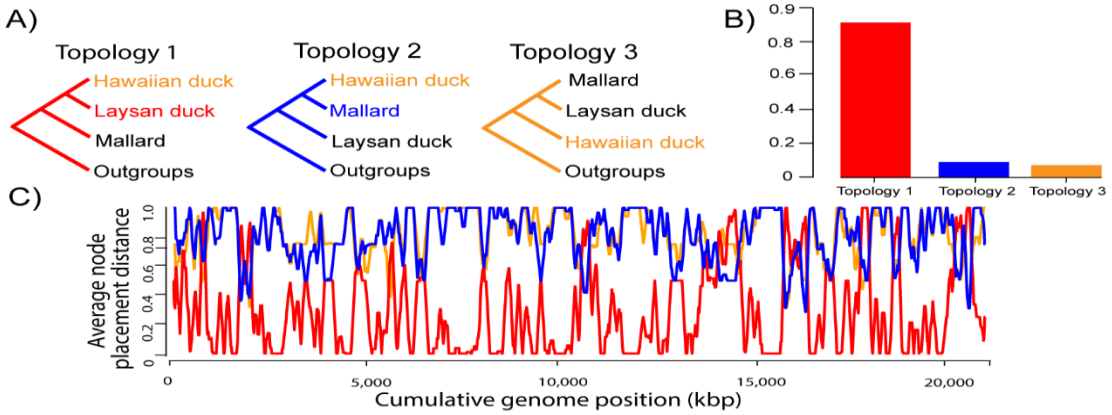

Figure S1.3. Phylogenetic inference across the Z-chromosome to assess the origin of the Hawaiian duck. A) The main three alternative topologies for the position of the Hawaiian duck, and B) their relative weightings across the Z-chromosome. C) Average node placement distance across the Z-chromosome.

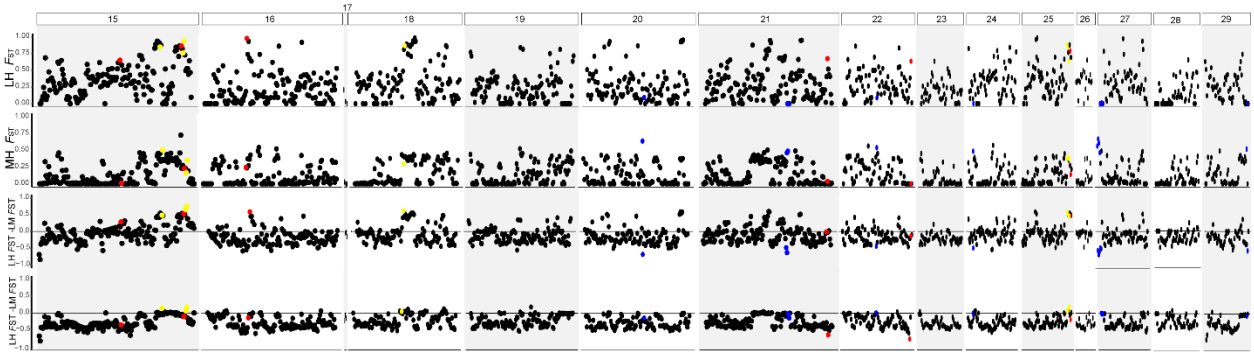

Figure S1.4. Divergence peaks between the Hawaiian duck and parent taxa on microchromosomes. F_{ST} plots for the LH (top), and MH (second from top) for 100kb non-overlapping windows across the genome. The bottom two panels display the disparities in FST values between each LH and MH comparison from the parent species comparison for the same 100kb non-overlapping windows. Windows identified as LH divergence peaks are shown in red, MH in blue, and PH peaks in yellow.

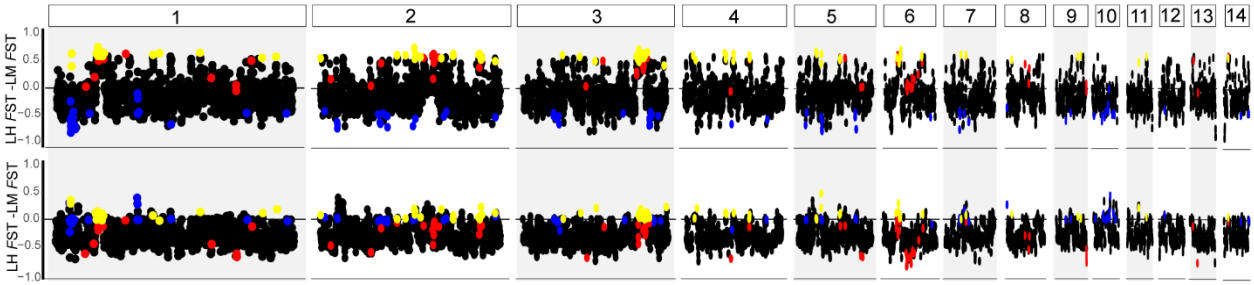

Figure S1.5. Divergence peaks for PH regions on large chromosomes. F_{ST} disparities between the LH and LM comparisons for 100kb non-overlapping windows across the large chromosomes (to) and disparities between MH and LM comparisons in the bottom panel. LH divergence peaks are shown in red, MH in blue, PH in yellow.

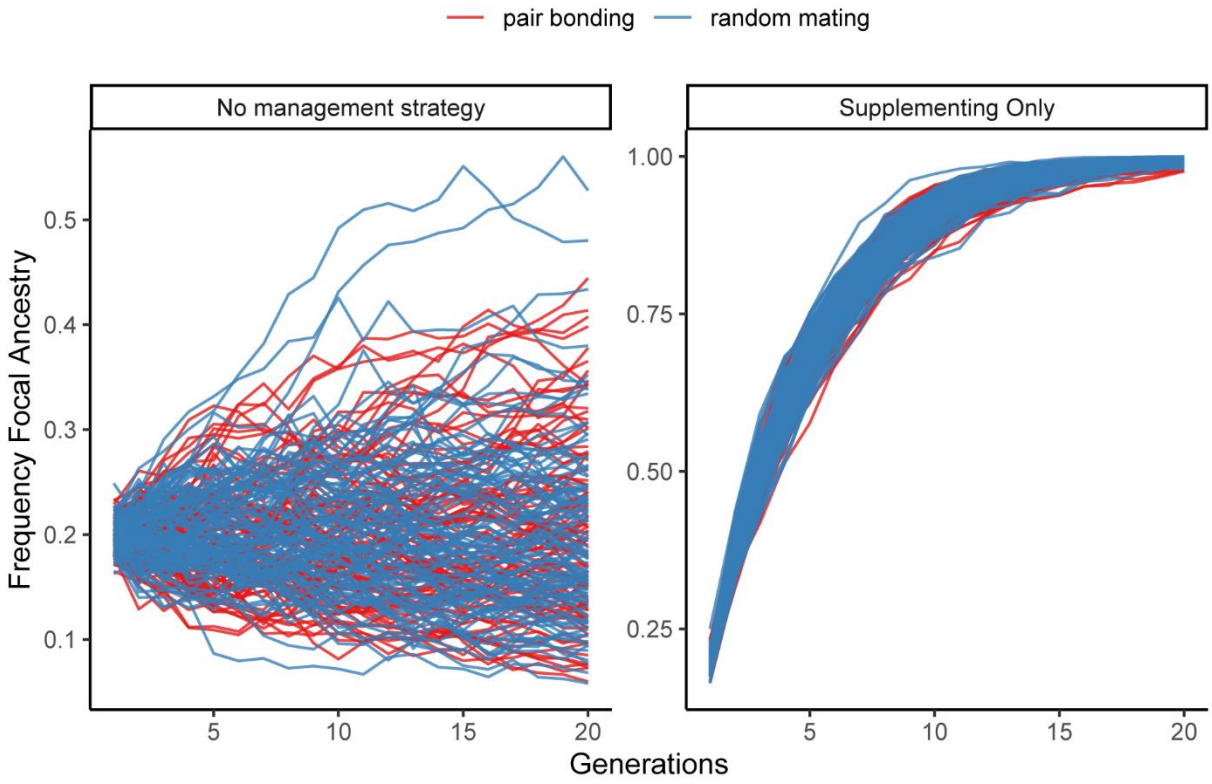

Figure S2.1. Comparing outputs for pair bonding (red) and random mating (blue) mating methods based on a (A) no management strategy versus (B) optimization of supplementation only, each run 100 times. As expected, no substantial change between the starting and final ancestries were attained with either mating method if no management is done (Wilcox-test p-value = 0.32), whereas statistically similar results of substantial ancestry improvement was recovered with both mating methods when optimizing supplementation (Wilcox-test p-value = 0.73).

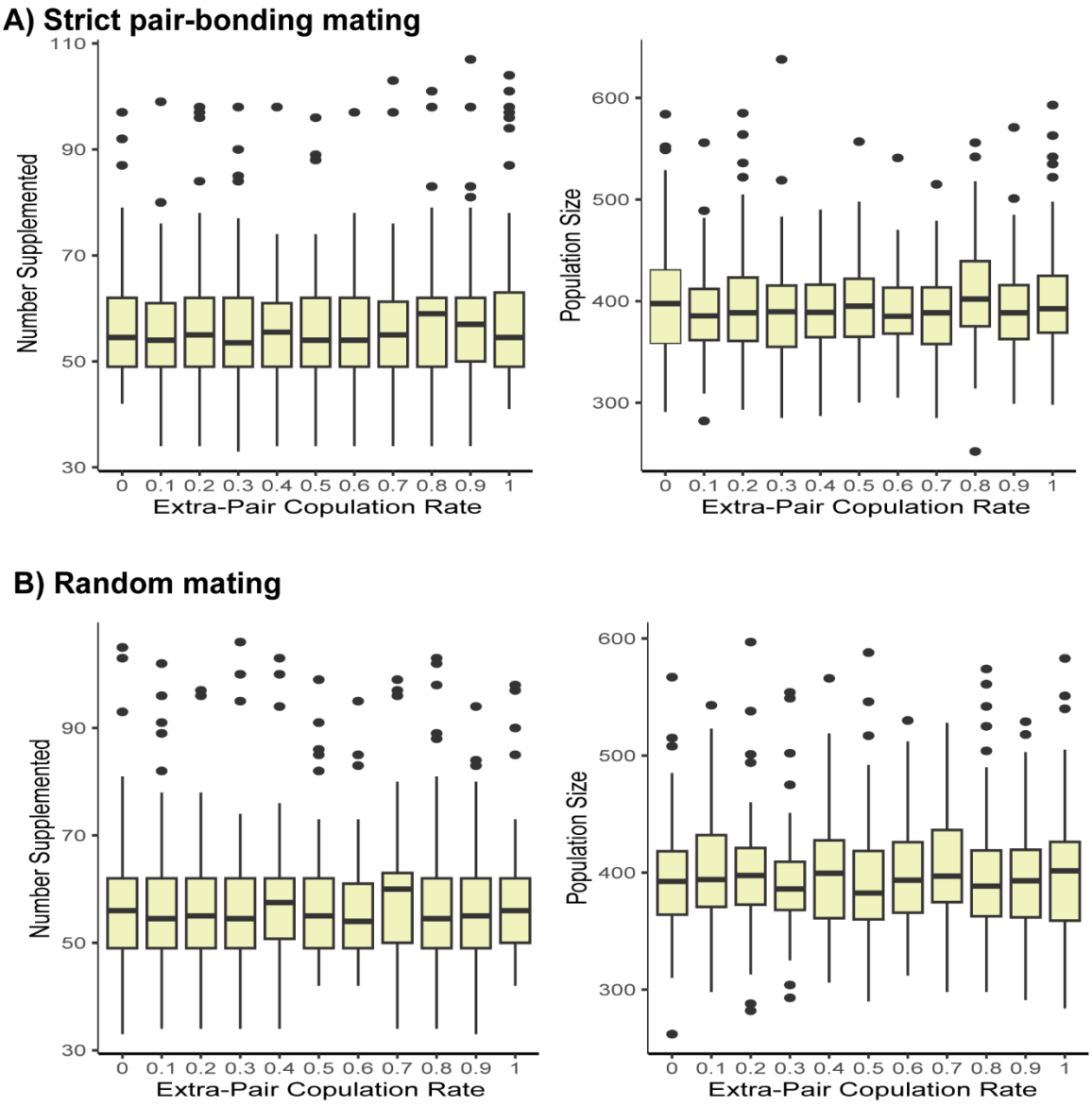

Figure S2.2. Changes in optimized supplementation and final population sizes when varying extra-pair copulation rate. Comparing these changes when using a (A) Strict pair-bonding or a (B) Random mating model.

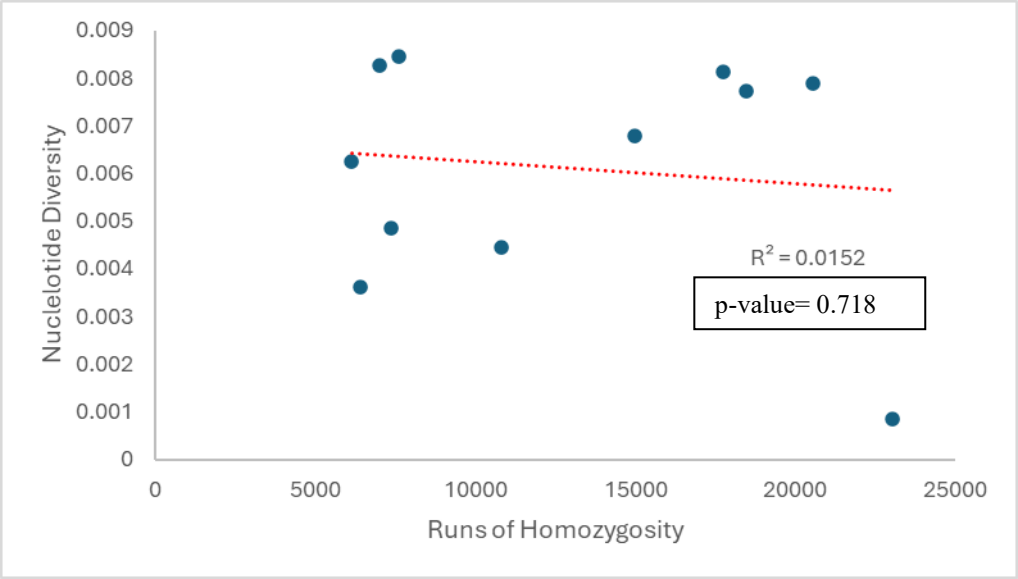

Figure S3.1. Regression analysis between Nucleotide diversity and runs of homozygosity across the species of the Mallard Complex.

Vita

Flor Brigitte Hernández Camacho was born in Lima, Peru. Her parents are Flora Camacho and Elvis Hernández. She received her elementary education at San Judas Tadeo and her secondary education at Alfa in San Juan de Lurigancho. In March 2006, she enrolled in the Science College at Universidad Nacional Agraria La Molina (UNALM) and graduated with honors with a B.S. in Biology in December 2010. In 2008, she joined the Centro de Ornitologia y Biodiversidad (CORBIDI) as a volunteer, later becoming an associated researcher in 2011. She was admitted to the Graduate School at the University of Texas at El Paso in January 2018, and graduated in the summer of 2024. Subsequently, she relocated to Ithaca, New York, with her husband and son, Alexis and Ricardo Díaz, where she worked as a postdoctoral researcher at the Cornell University's Lab of Ornithology.

Contact Information: fhernandez@corbidi.org