GENOMIC SIGNATURES OF POPULATION HISTORY IN A PAIR OF RECENTLY DIVERGED AUSTRALIAN TEAL SUPPORT STRONG SELECTION ON THE Z-SEX CHROMOSOME

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

KEVIN K. HAWKINS
B.A., Miami University, 2014

2016
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

September 17, 2012

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Kevin K. Hawkins ENTITLED Genomic signatures of population history in a pair of recently diverged Australian teal support strong selection on the Z sex-chromosome BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

_____

Jeffrey L. Peters, Ph.D.
Thesis Director

_____

David Goldstein, Ph.D.
Chair, Department of Biological Sciences

Committee on
Final Examination

_____

Jeffrey L. Peters, Ph.D.

_____

Scott E. Baird, Ph.D.

_____

John O. Stireman, Ph.D.

_____

Robert E.W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

ABSTRACT

Hawkins, Kevin K. M.S. Department of Biological Sciences, Wright State University, 2016. Genomic signatures of population history in a pair of recently diverged Australian teal support strong selection on the Z sex-chromosome.

Sex chromosomes are thought to be an important component of the genome associated with speciation and the buildup of reproductive isolation. Recent advances in sequencing technologies and improvements in population genetics and modeling techniques have made it possible to better assess genomic signatures of selection, genetic drift and gene flow in diverging lineages. Recent studies have shown elevated differentiation on the Z sex-chromosome between the Australian grey teal (*Anas gracilis*) and chestnut teal (*Anas castanea*). Here, we used next generation sequencing to scan ~3,400 autosomal loci and ~190 Z loci to examine genomic differentiation and signatures of selection and gene flow between grey and chestnut teals. We also inferred demographic history to assess gene flow and signatures of selection. We found weak differentiation in autosomal loci (mean Φ_{ST} = 0.008), but ~ 28-times higher differentiation on the Z-chromosome (mean Φ_{ST} =0.23). We also found that this higher differentiation was localized on the q-arm of the Z chromosome between 15-million and 40-million base pairs. Whereas we could not reject a strict isolation model on the Z chromosome, models that incorporated gene flow provided a significantly better fit for autosomal loci, which is consistent with reduced local gene flow as expected under a speciation with gene flow model. However, nucleotide diversity within chestnut teal was

reduced within the region of elevated differentiation on the Z chromosome, which is more consistent with a selective sweep rather than reduced effective gene flow. Furthermore, linkage disequilibrium within this region was elevated with respects to other regions on the Z chromosome as well as more significantly elevated in chestnut teal than in grey teal. Although we cannot fully reject a model of speciation with gene flow, recent speciation seems a more plausible explanation of the nearly absent divergence of mitochondrial and autosomal DNA. Rather, these data suggest that a selective sweep coupled with strong linkage disequilibrium on the Z chromosome might have played a role in speciation.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jeff Peters, for his unwavering support, mentorship, and invaluable advice through the project presented here. I thank Dr. Leo Joseph and Dr. David Roshier for their loan of tissue samples, and I thank Dr. John Stireman and Dr. Scott Baird for their service on my thesis committee, advice, and guidance. I also thank Dr. Philip Lavretsky and Joel Nelson for their help with lab protocols, data analysis, and intellectual support during this project. Lastly, I dedicate this thesis to my son, Aayden Alexander Hawkins, and I thank my wife, HeavinLee Hawkins, for her continued support and encouragement.

INTRODUCTION

Speciation occurs along a continuum from incipient stages of divergence to subspecies to fully reproductively isolated species (Mallet, 2008; Nosil *et al.*, 2009). A recent focus of evolutionary biology is to use genomic analyses to study the genetic signatures of this process to generate hypotheses about mechanisms of speciation. Whereas there are some expectations for signatures expected under different models of speciation, these expectations are constantly being revised with new data and improved modeling techniques. For example genetic drift, which is the random sampling of gametes during sexual reproduction, reduces diversity within a population and causes population divergence. Selection can drive divergence when there is an association between genotype and fitness whereby either advantageous alleles become fixed (while deleterious ones are removed) or allelic diversity is maintained. Selection also appears to have a more substantial effect in a region of low recombination. Selection in areas of high linkage disequilibrium can lead to hitchhiking which pulls multiple adjacent loci along to fixation (Smith and Haigh, 1974; Charlesworth, 2008). While these adjacent loci are not the targets of selection directly, they will exhibit the same or very similar genomic signatures of selection. When hitchhiking occurs, we expect to find large regions beyond an individual locus differentiated that has implication in gene flow and introgression (Nosil & Feder, 2013; Via, 2012). It is these large regions that become genomic islands of differentiation between two species.

A growing body of literature suggests some species undergo divergent hitchhiking (divergent selection in areas of high linkage disequilibrium), which results in regions or islands of divergence that result from reduced effective gene flow (Via, 2009; Feder &

Nosil, 2010; Ellegren *et al.*, 2012; Renaut *et al.*, 2013; Hemmer-Hansen *et al.*, 2013; DeFarveri *et al.*, 2013). Genomic island formation in populations with gene flow is thought to occur at particular loci in the genome that experience strong enough linked selection that is sufficient to overpower the homogenizing effect of gene flow (Michel *et al.*, 2010; Nadeai *et al.*, 2011; Hemmer-Hansen *et al.*, 2013). This concept however, raises a question about recombination rates. In regions of high recombination, the hitchhiking effect can be inhibited due to disassociation of linked loci. A theoretical approach using simulated data was used to explore the effects of recombination in the islands of divergence model (Feder and Nosil, 2010) and it concluded that initial island formation is dependent on regions of reduced recombination (Feder and Nosil, 2010). As selection acts on a region with high linkage disequilibrium (and reduced recombination), adjacent loci hitchhike along with the locus under selection (Charlesworth *et al.* 1997; Via, 2009) which creates an island that has limited gene flow with respects to the background levels of the genome (Feder and Nosil, 2010). In a recent empirical study, a genomic scan using four divergent pairs of *Helianthus* sunflower species showed a tendency towards reduced recombination locally in genomic islands relative to the rest of the genome (Renaut *et al.*, 2013).

An alternative hypothesis for the formation of islands was presented by Cruickshank and Hahn (2014), who suggested that reduced genomic diversity can result from post speciation selection, not reduced effective gene flow, and this causes the formation of islands. Most studies suggesting a speciation with gene flow model use various derivatives of the fixation index (F_{ST} or when using sequence data Φ_{ST}), which is a measure of the amount of differentiation between two populations. Cruickshank and

Hahn (2014) argue that the sole reliance on the fixation index could be misleading because fixation indices measure differentiation relative to within population diversity. This becomes important because linked selection (both selection in favor of advantageous mutations and selection against deleterious ones) reduces within-population diversity in the region of selection relative to the rest of the genome. A lack of recombination, thereby maintaining the association of linked loci, causes larger regions to exhibit reduced genetic diversity (Charlesworth, 1998, Renaut *et al.*, 2013). These large regions of reduced nucleotide diversity cause those same regions to exhibit elevated relative divergence, regardless of the interaction between selection and gene flow. In contrast, Curickshank and Hahn (2014) predict that under models of reduced gene flow within islands (speciation with gene flow), absolute divergence, which is less affected by within-population diversity, should be elevated relative to background levels of divergence throughout the genome.

Reanalysis of empirical data from *Ficedula* flycatchers suggests that the strong signatures of islands of divergence, that were initially reported as evidence of divergence with gene flow (Ellegren *et al.*, 2012), were not evident when examining absolute divergence (Cruickshank & Hahn, 2014). In fact, the absolute divergence was reduced in comparison to background divergence, and the high relative divergence resulted from low nucleotide diversity within the islands. Given the strong signatures of selection, low recombination, and high linkage disequilibrium, these results are consistent with a model of selection without gene flow (Cruickshank & Hahn, 2014). Supple *et al.* (2015) further empirically explored islands in *Heliconius* butterflies and concluded that they are caused by strong selection, rather than differential gene flow, on wing color pattern. It is

important to note that there is a distinctive difference between post speciation selection which exhibits high relative divergence due to a within population process of linked selection and speciation with gene flow which is a between population process whereby selection limits between population recombination at sites of ecological relevance (Via, 2012).

Sex chromosomes provide an interesting case for the study of speciation for a variety of reasons. In particular, the sex chromosomes appear to diverge at a faster rate than autosomes. For example in the heterogametic sex (e.g., males in mammals and flies, XY; females in birds and butterflies, ZW), is most affected by hybrid inviability or sterility (Haldane, 1922). Furthermore, heterogametic hybrids are more affected by hybrid sterility regardless of whether the alleles involved in hybrid compatibility, or lack thereof, were dominant or recessive, whereas homogametic hybrids (e.g., one copy of the Z chromosome from each parent species) are only affected when the alleles associated with hybrid incompatibility are dominant (Turelli & Orr,1995).

Furthermore, the X and Z chromosomes have about ¾ the effective population size of the autosomes, and therefore genetic drift will be more effective in randomly fixing alleles throughout the sex chromosomes in comparison to the autosomes. The greater efficacy of drift should cause more rapid divergence on the sex chromosome relative to the autosomes (Ohta & Gilespie 1996; Mank *et al.*, 2009; Johnson & Lachance, 2012). Genetic drift is also a more effective cause of genetic divergence in ZW systems, in which females are the heterogametic sex, because sexual selection in males (i.e., high variance in reproductive success) would cause the effective population size of the Z to be even less than ¾ that of autosomes (Mank *et al.* 2010). Sex chromosomes

have been implicated in birds for plumage characteristics, male fitness, and female mate preference, and might contain targets of strong sexual selection (Saetre *et al.* 2007; Mank, 2009; Pryke, 2010; Huang & Rabosky, 2015).

Study system

In this study we analyzed genomic divergence in a pair of morphologically, and behaviorally distinctive duck species. The monochromatic grey teal (*Anas gracilis*) is morphologically characterized by light dull-brown plumage and a whitish throat in both sexes. The dichromatic chestnut teal (*Anas castanea*), unlike the grey teal, are characterized by males that exhibit a vibrant green head and chestnut body (Marchant & Higgins, 1990), whereas females resemble grey teals with a slightly darker color on the throat (Marchant and Higgins, 1990). Geographically, the grey teal has a relatively high dispersal capability and is found throughout Australia and parts of New Guinea and Indonesia (Marchant and Higgins, 1990). In contrast, the chestnut teal, seems to be more sedentary and is restricted to southeastern and southwestern regions of Australia (Marchant & Higgins, 1990). These two species hybridize in captivity, but it is unclear whether gene flow occurs in the wild (Marchant & Higgins, 1990).

Previous genetic work has shown that grey and chestnut teals are not significantly differentiated at mitochondrial DNA (mtDNA) or autosomal loci (Joseph *et al.* 2009; Dhami *et al.* 2013, 2016). However, relative divergence (i.e., Φ_{ST}) was 50-times higher on the Z chromosome compared to autosomes (Dhami *et al.* 2016). Examining only seven Z-loci, Dhami *et al.* (2016) described a potential region of elevated divergence near the center of the Z chromosome. Within this region, chestnut teal exhibited a high frequency of derived alleles, which suggests that selection within the dichromatic chestnut teal might have contributed to the high divergence.

Dhami *et al.* (2016) could not reject either a model of speciation with gene flow or a model of post-speciation selection. Based on coalescent simulations, the authors described elevated absolute divergence in Z loci within the region of elevated relative divergence, which is consistent with a speciation with gene flow model. However, they found no significant difference in absolute divergence between Z loci and autosomal loci but rather found reduced nucleotide diversity on the Z chromosome, which is more consistent with a post-speciation selection model as described by Cruickshank and Hahn (2014). Furthermore, the density of markers was insufficient to determine whether elevated differentiation between grey and chestnut teal characterized the entire Z chromosome or whether there was an island of differentiation within the chromosome. More data are needed to distinguish between these hypotheses (Nosil and Feder, 2012).

In this study we expanded upon Dhami *et al.* (2016) using double-digest restriction-associated-DNA sequencing (ddRAD-seq; DaCosta and Sorenson, 2014) to scan the genomes of two recently diverged Australian teals. Our objectives were to determine whether elevated divergence on the Z-chromosome is associated with an island of divergence and to further test models of speciation. We examined patterns of divergence across ~3,300 autosomal markers and 194 Z-linked loci. In addition, for loci with elevated divergence (outlier loci), we tested for evidence of selection acting within one or both species to help address which model is more consistent with the patterns of divergence of the Australian teals.

MATERIALS AND METHODS

Sampling

We sampled a total of 39 Australian teals, including 21 chestnut teal and 19 grey teal; each species was sampled from widely distributed locations throughout their respective ranges (see Dhami *et al.*, 2016). DNA was extracted from blood or tissue samples using a DNeasy Blood and Tissue kit following manufacturer protocols (Qiagen, Valencia, CA). We measured the concentration of DNA using a NanoDrop 2000 Spectrophometer to ensure a concentration of at least 20 ng/µl for each sample.

Double digested RAD (ddRAD-seq) library prep

Following DaCosta and Sorenson (2014), we prepared ddRAD-seq libraries by digesting ~1 ug of genomic DNA using 10 Units of restriction enzymes SbfI and EcoRI. We ligated Illumina adapters, containing barcodes for de-multiplexing, to the sticky ends generated by the restriction enzymes. We then used a 2% low-melt agarose gel and a MinElute gel extraction kit (Qiagen, Valencia, CA) to select DNA fragments between 300 and 450bp. We then used PCR with Phusion-HF DNA polymerase (Thermo Scientific, Pittsburgh, PA) to amplify size-selected fragments. After amplification, fragments were cleaned using magnetic AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN) and quantified using quantitative PCR and a KAPA Library Quantification Kit (KAPA Biosystems, Inc., Wilmington, MA). Samples were pooled in equimolar concentrations, and multiplexed libraries were sequenced on an Illumina HiSeq 2000 at Tufts University.

Quality Control and Filtering

De-multiplexing, filtering, and assembling Illumina reads were done using a custom python pipeline described in DaCosta and Sorenson (2014) (available at http://github.com/BU-RAD-seq/ddRAD-seq-Pipeline). In brief, we assigned reads to individuals based on unique combinations of ligated barcodes and indices. We filtered out loci with low reads (reads with average quality score <30) and collapsed identical reads while retaining read count and the highest quality scores. We used the UCLUST function in USEARCH v. 5 (Edgar 2010) to cluster reads that were <10% divergent. Each locus was then mapped to the mallard reference genome (Kraus et al., 2011; Huang et al., 2013) using BLASTN v.2 (Altschul *et al.*, 1990); loci that mapped to the same position within the mallard genome were subsequently combined. Reads within a cluster were then aligned using MUSCLE v.5 (Edgar, 2004). Individuals were genotyped as follows based on DaCosta and Sorenson (2014); we designated an individual homozygous if > 93% of reads were identical and heterozygous if a second allele was observed at a frequency of at least 29%. Individuals were also scored as heterozygous if a second allele was represented by 20-29% of reads and the second allele was found in other sampled individuals. For individual genotypes that failed to meet either of these criteria or had evidence of greater than two haplotypes (a third allele was represented by >7% of reads), we retained the allele represented by the majority of reads and scored the second allele as missing data. Likewise, a single allele was scored for loci containing fewer than ten reads (low depth). We retained all loci that were genotyped for $\geq$90% of individuals and flagged in <5% of individuals, for a total of 3,530 loci.

Putative loci with end gaps due to an insertion/deletion or a polymorphism in the SbfI restriction site were either trimmed or flagged for manual editing. Also, loci with > 2

polymorphisms in the last five base pairs or >19 segregating sites overall were flagged for manual inspection. Manual inspection and editing for 353 loci was done using Geneious (Biomatters Inc., San Fransicsco, CA).

To determine the physical location of each locus on chromosomes, we conducted a BLAST search against the assembled mallard reference genome (obtained from T. Faraut; unpubl. data) using the search sequence function in Geneious. In cases in which we obtained multiple hits, we retained hits that had >90% sequence identity and >90% query coverage; loci that had multiple hits with high identity and coverage were not mapped to chromosomes. In the event that sequences did not BLAST to the reference genome we classified loci as either autosomal or Z by two different methods. The first method was to BLAST to the chicken (*Gallus gallus*) reference genome (accession numbers PRJNA10807-08, PRJNA13342, PRJNA202483), and any additional loci the matched to the chicken Z or an autosome were classified accordingly. For all other remaining loci, we used the approach outlined in Lavretsky *et al*. (2015) in which depth and homozygosity across markers were examined. Lavretsky *et al*. (2015) reasoned that since females are heterogametic for the Z chromosome (i.e., females are hemizygous), Z loci should have no heterozygosity in females and about one-half the depth of males. In contrast, heterozygosity and depth should be similar between females and males for autosomal loci. Based on loci that had BLAST hits to mallard autosomes or the Z-chromosome, we categorized loci with a relative depth (male depth/female depth) of > 2.2 as Z-linked and < 2.2 as autosomal loci. In total, 3,162 loci were classified as autosomal based on BLAST hits to the duck (3,162 loci) and chicken genome (17 loci) or

based on relative depth (157 loci), and 194 loci were classified as Z-linked based on hits to the duck (173 loci) and relative depth (21 loci).

Diversity, divergence and population structure

We calculated overall intraspecific nucleotide diversity (π; the average number of pairwise differences among sequences for each species), absolute divergence (d_{xy}; the average number of differences between sequences from different species), and relative divergence (Φ_{ST}; the proportion of nucleotide differences partitioned between species) using the PopGenome package for R (Pfeifer *et al.*, 2014). For intraspecific diversity and absolute divergence, we divided the values obtained (average number of differences per locus) by the length of each locus to standardize values to the number of differences per site. We then generated a map of relative divergence, absolute divergence, and intraspecific diversity across the genome based on the BLAST hits to the mallard reference genome (see above). We used a t-test to compare diversity and absolute divergence between regions of elevated relative divergence. We further used a Mann-Whitney U test (two-tailed) to compare nucleotide diversity between grey and chestnut teal (as done in Chapman *et al.*, 2016). Significance was determined at $P \leq 0.05$.

We also examined overall population structure and our ability to assign individuals to species using the program ADMIXTURE (Alexander *et al.*, 2009). ADMIXTURE uses a maximum likelihood approach to assign an individual to a specific population defined a priori. We tested models of population structure that included 1–5 populations. We included a cross validation method to determine the number of populations that received the highest support. We treated autosomal and Z-linked loci in separate analyses.

We also used SPLITSTREE 4.14 (Hudson and Bryant, 2006) to construct neighbor-net networks to assess the genetic distances between individuals. We concatenated loci for autosomes and the Z-chromosome separately and used a custom python script to obtain consensus sequences for individuals using IUPAC ambiguity codes for heterozygous sites. Networks were constructed using uncorrected *P*-distances and average states for polymorphic positions (i.e., C vs. T equals a distance of 1.0, whereas C vs. Y equals a distance of 0.5).

Outlier loci

To determine whether loci significantly deviated from background levels of differentiation, we used BayeScan v. 2.01 (Foll and Gaggiotti, 2008), which uses MCMC to assign a probability of each locus conforming to a model of selection or neutrality based on its Φ_{ST} in relation to all other loci in the dataset. We ran 20 pilot runs of 5,000 steps and a final run with a burn-in of 50,000 steps and 5,000 sampling steps using a thinning interval of 10 steps.

For loci linked to the Z-chromosome, we determined the frequency of derived SNPS for grey teal and chestnut teal. We determined ancestral and derived SNPs by comparing sequences to ddRAD-seq loci obtained from the Madagascar teal (*Anas bernieri*) (Peters, unpubl. data), which is the closest phylogenetic relative of the Australasian teals (Johnson and Sorenson, 1999), and to the mallard genome. In our dataset, if the SNP had the same nucleotide as Madagascar teal and mallard it was considered to be the ancestral state. If the SNP differed from both the mallard and the Madagascar teal, we considered the SNP to be derived. We also considered any SNP that was inconsistent between our data set and the mallard (in the event of missing data in the

Madagascar teal) to be derived. All SNPs that could not be matched to either reference database or differed between the Madagascar teal and the mallard were excluded from this analysis.

We assessed the degree of linkage disequilibrium using Haploview (Barrett *et al.*, 2005). We independently analyzed chestnut teal and grey teal Z chromosome data. We also categorized our data into three categories which was based on location on the Z chromosome (5' end, island region, and 3' end of the Z chromosome). We utilized r^2 (Hill and Robertson 1968) to assess the degree of linkage decay. r^2 is commonly used to assess the degree association between two SNPs (two SNPs completely associated with each other—complete linkage disequilibrium), would have an r^2 of 1 whereas two SNPs that are independent (linkage equilibrium) will have an r^2 closer to 0. We only used parsimony informative polymorphic sites for this analysis (only biallelic SNPs in which the minor allele was present at least twice). We were particularly interested in the levels of and rate of linkage decay between chestnut teal and grey teal across the Z chromosome.

Demographic history

To determine whether signatures of population demography, including estimates of gene flow and divergence times, differed between autosomes and the Z-chromosome, we used the Diffusion Approximation for Demographic Inferences (∂a∂I; Gutenkunst, 2009). ∂a∂I has been used in several studies to test demographic scenarios including models of migration/no migration and population size changes (Gutenkunst, 2009; McCoy *et al.*, 2014; Liu *et al.*, 2014). ∂a∂I uses diffusion approximation to model the site frequency spectrum (SFS) under various models of divergence, and attempts to minimize

the residuals and thereby maximize the similarity between the modeled SFS and the empirical SFS. We used a folded spectrum (to account for a lack of ancestral information) of our data to compare models with and without gene flow. Autosomal and Z-linked loci were analyzed separately. We also analyzed the region of elevated divergence separately from the rest of the Z chromosome.

We used a concatenated NEXUS alignment of the autosomal and Z datasets to obtain the frequency spectrum (modified from http://github.com/mgharvey/misc_python). We analyzed the data using a standard isolation-with migration (IM) model and a strict isolation (I) model. We ran multiple (~10) independent runs of ∂a∂I to confirm parameter convergence on similar estimated values. We then did a parametric bootstrap using the Godambe Information Matrix (GIM) function of ∂a∂I to obtain parameter standard deviations from 100 random re-samplings. We also used an adjusted log-likelihood ratio test (LRT) using the general model (IM) and the nested model (Isolation) to determine if a migration or no migration model provided a better fit to the data (Coffman *et al.*, 2016). Since ∂a∂I assumes each SNP is independent we accounted for linkage disequilibrium by using the GIM as explained in Coffman *et al.* (2016).

Here we adopted a strategy similar to Liu *et al.* (2014) in which we did multiple runs of ∂a∂I, under different models to determine if the models fit to the data significantly improved by adding a flexible parameter. We were particularly interested in determining if there was a significant difference between a strict isolation model and a model that included a flexible migration parameter. In this analysis we were less focused on actual estimates of parameters, as these estimates vary between models, and more focused on fitting the model to the data. We note that care must be taken when assessing parameter

values as with many demographic inference programs, because positive/linked selection (selective sweep) can lead to drastically overestimated parameters, especially if using a small SNP dataset. Recent work, however, has shown that ∂a∂I is relatively robust to selection with regard to other demographic inference programs (Schrider *et al.*, 2016).

We expect that if the speciation with gene flow model is supported, then we should detect evidence of gene flow within regions of low divergence. In contrast, an isolation model or a model with substantially lower rates of gene flow should be a better fit to regions of elevated divergence (Via, 2012).

RESULTS

Loci recovery and chromosomal locations

After quality filters, we retained 3,530 total loci (449,537 bp) that were recovered from >90% of individuals and contained ≤10% flagged genotypes. We obtained on average 480,000 (±302,000 S.D.) reads per individual and a median read depth of 82 reads (Range = 0 - 1,324). We categorized 3,336 loci as autosomal (424,812 bp; including 3,101 variable loci) and 194 Z loci (24,725 bp; including 157 variable loci). Of these loci, we mapped 173 Z and 3,162 autosomal loci to the mallard chromosomes.

Diversity, divergence and population structure

Analyses of population structure suggest a discordance between autosomal and Z loci (Fig. 1). In the autosomal neighbor net tree, chestnut teal and grey teal are intermixed (Fig. 1A)—some grey teal individuals were more similar to chestnut teals than to other grey teals and vice versa. In contrast to the autosomal network, the Z neighbor net tree shows a clear genetic split between grey and chestnut teals (Fig. 1B). Likewise, ADMIXTURE analyses revealed a lack of detectable differences between

species for autosomal DNA (Fig. 2B), whereas the data for the Z chromosome was consistent with a two-population model (K=2) that recovered grey and chestnut teal as being distinct (Fig. 2B). Most individuals were assigned to their respective species with >99.9% posterior probability, although three grey teal had some evidence of admixture from chestnut teal.

In measures of relative divergence we found that autosomal loci were weakly differentiated (composite Φ_{ST} = 0.0083), and BayeScan results suggested that differentiation at one locus (Φ_{ST} = 0.11; P<0.05) was significantly elevated compared to background levels (Fig. 3A). Differentiation at four additional autosomal loci was also somewhat elevated, but they were not significant outliers. The Z chromosome exhibited about 27 times greater interspecific differentiation (composite Z Φ_{ST} = 0.23) in comparison to the autosomes (Fig. 3). We found two loci (~1%) that had significantly elevated differentiation from the background Z differentiation and another four loci with a strong trend suggesting elevated differentiation (Figure 3B). All loci with significantly or nearly-significantly elevated differentiation clustered between 15 million and 40 million base pairs on the Z chromosome. Analyzing autosomal and Z loci together, differentiation at 1 autosomal and 18 Z loci were significantly elevated.

Mapping loci to chromosomal position, we found a broad region of elevated differentiation between 15 million and 40 million base pairs on the Z chromosome. This region included all 18 Z loci detected as significant outliers and had approximately 16 times greater relative divergence compared to the rest of the Z chromosome (mean Φ_{ST} = 0.282 vs. Φ_{ST} = 0.017, respectively; P < 0.0001). The autosomal outlier locus mapped to chromosome four.

In contrast to relative divergence, absolute divergence did not indicate a noticeable region of elevated divergence on the Z chromosome (figure 4A). Indeed, d_{xy} within the region of elevated relative divergence was only slightly elevated (mean $d_{xy(island)} = 0.041$; mean $d_{xy(non\text{-}island)} = 0.035$); however, this difference was not significant ($P = 0.06$). Z chromosome genetic diversity was approximately 1.3 times greater in grey teal than in chestnut teal (composite $\pi_{GRTE} = 0.0034$, composite $\pi_{CHTE} = 0.0026$; figure 4C and 4B). Genetic diversity in chestnut teal was especially reduced within the region of elevated relative differentiation; nucleotide diversity was five times higher outside this region (Fig. 4B), and this difference was significant ($P = 0.0001$). Grey teal did not exhibit reduced genetic diversity within the region of elevated differentiation, but rather diversity was more uniform across the Z-chromosome (Fig. 4D; $P= 0.35$). Furthermore, nucleotide diversity was significantly higher in grey teal than chestnut teal within the region of elevated relative differentiation ($P < 0.0001$), but not outside this region ($P = 0.30$).

Weak absolute differentiation was similarly found in the autosomal markers (Fig. 5A). Absolute divergence was about two-times higher for autosomal loci ($d_{xy} = 0.0067$) than to Z-loci ($d_{xy} = 0.0039$). Autosomal nucleotide diversity was nearly indistinguishable between chestnut and grey teal ($\pi_{CHTE} = 0.0066$ and $\pi_{GRTE} = 0.0067$; $P = 0.99$; Fig. 5B and Fig. 5C).

On the Z chromosome, the ancestral state could be determined for 83 SNPs within the region of elevated differentiation and 150 SNPs outside this region. Within the elevated region, 69 SNPs were fixed in the chestnut teal—58 were fixed for the ancestral variant and 11 were fixed for the derived variant (Fig. 6A). Note that within this region

for chestnut teal, the frequency of derived variants was either $\leq$0.20 or $\geq$0.75; derived SNPs at intermediate frequencies were lacking (with the exception of one locus at derived frequency of 0.40). In contrast, in the same region for grey teal, only 20 SNPs were fixed, 16 of which were fixed for the ancestral variant (Fig. 6A, blue), and there were nine derived SNPs with intermediate frequencies (>0.2 and <0.75). In three of the four instances where differences were fixed between chestnut and grey teal (F_{ST} = 1.0), the derived variant was fixed for the chestnut teal. In the final instance of this, the grey teal was fixed for the derived variant. Markers that fell outside the region of elevated differentiation exhibited a similar distribution of derived SNP frequencies between grey and chestnut teal with the exception that there were more fixed ancestral SNPs in chestnut teal (Fig. 6B). Furthermore, several derived variants were at intermediate frequencies for both species. In fact the number of intermediate SNPs were exactly the same in chestnut teal and grey teal (14 Intermediate SNPs and 136 fixed SNPs in both chestnut teal and grey teal).

Examining overall linkage disequilibrium using r^2 for chestnut teal, we found substantially higher linkage within the region of elevated divergence compared to other regions of the Z chromosome (Fig. 7). The 3' end of the Z chromosome had slightly higher linkage disequilibrium compared to the 5' end (Fig. 7). In contrast, grey teal exhibited similar levels of linkage on the Z chromosome across all three regions of the Z chromosome (e.g., 3' end, 5' end, and the elevated region). Whereas linkage across the 3'- and 5'-ends were similar between chestnut and grey teal, linkage was substantially higher for chestnut teal within the region of elevated divergence.

Demographic history

Estimating demographic histories using the site frequency spectrum in ∂a∂i, we found the model that included migration provided a significantly better fit than a strict isolation model for autosomal loci (P=0.005). However, we could not reject a strict isolation model for all Z loci ($P = 0.07$), Z loci outside the region of elevated differentiation (P=0.18), and Z loci within the island of divergence (P=0.165).

DISCUSSION

Sequences from 3,530 ddRAD-seq loci collected from two species of Australasian teal confirm and expand on the findings of Dhami *et al.* (2016) that differentiation on the Z chromosome is elevated relative to autosomal DNA. We found nearly indistinguishable differentiation between the species on the autosomal markers, yet on the Z chromosome we found that differentiation was 28 times higher (Fig. 3) and more consistent with expectations for separate species (Figs. 1 & 2). Dhami *et al.* (2016) reported a 50 times elevated relative divergence using seven Z-linked and 17 autosomal loci. The difference between our results and Dhami *et al.*'s (2016) results is likely attributable to our more comprehensive sampling that provided a much broader distribution of markers across the Z chromosome. Furthermore, Dhami *et al.* (2016) suggested the possibility of an island of divergence on the q-arm of the Z-chromosome. However, they only examined seven Z loci, which was insufficient coverage to unambiguously demonstrate the presence of an island. Here, using 194 loci scattered across the chromosome, we documented a large genomic island located between 15-million and 40-million bp on the q-arm of the chromosome. All three loci with elevated differentiation found by Dhami *et al.* (2016) fell within this same region.

The genomic evidence strongly suggests neutral genetic drift is unlikely to be the sole cause or primary reason for elevated differentiation. First, drift alone is unlikely to explain the Z:A ratio for Φ_{ST} between chestnut and grey teal of ~27.7. Using simulated data, Lavretsky *et al.* (2015) found that a Z: A ratio of this magnitude could only be explained by neutrality if the Z:A ratio for effective population size was <0.1. Second, if genetic drift was the only evolutionary force acting on the Z chromosome then we would expect the overall diversity to be reduced in the population with the lower effective population size (chestnut teal) across both the Z chromosome and autosomal chromosomes. However, we find that genetic diversity is reduced in chestnut teal compared to grey teal only within the 15 million to 40 million bp region on the Z chromosome (Fig. 4A-D). In contrast, chestnut teal and grey teal have similar nucleotide diversity within the 3'- and 5'-ends of the Z chromosome and across autosomal loci. Third, two loci within the region of elevated differentiation were significant outliers (and three loci exhibited a strong trend) relative to background levels of differentiation on the Z-chromosome, and 18 loci were outliers relative to overall genomic background levels. Indeed, three loci exhibited fixed differences between grey and chestnut teal. Thus, there is a strong signature of selection acting within this region of the Z-chromosome.

Selection after speciation

A significant reduction in nucleotide diversity within the incidental island suggests that natural selection, either by positive selection fixing an advantageous allele and linked alleles, or purifying selection removing detrimental or slightly deleterious alleles from a population, is contributing to divergence (Maynard and Haigh, 1973; Charlesworth *et al.*, 1997; Messer & Petrov, 2003; Storz, 2005; Oleksyk *et al.*, 2009).

Interestingly, this signature of reduced nucleotide diversity within a specific region (15 million to 40 million bp) was only exhibited in chestnut teal (Fig. 4). Given a lack of this signature in grey teal, it is likely that selection is acting either on a few ecologically relevant loci with hitchhiking or on a large region of the Z chromosome in the chestnut teal.

A selective sweep or strong purifying selection in one population can result in a significant reduction in genetic diversity and thus increased relative divergence (Maynard and Haigh, 1973; Akey *et al.*, 2002) with a high frequency of fixed or nearly fixed alleles (Nielsen *et al.* 2005). We expect that under a model of post-speciation selection, there should be a high frequency of fixed SNPs in the species and region that are under selection. Within the island of differentiation compared to other regions on the Z chromosome, chestnut teal exhibited more loci that are fixed or nearly fix and lacked SNPs at intermediate frequencies. The contrast between these regions suggests that positive selection has driven advantageous alleles to fixation (Fig 6A; yellow). In contrast, grey teal have many intermediate allele frequencies with relatively fewer fixed or nearly fixed SNPs within this region (Fig. 6A; blue). Outside the island, the frequency of fixed alleles is similar between chestnut and grey teal (6B). These contrasts strongly support the conclusion that selection is acting independently in the chestnut teal.

In further support of selection in the chestnut teal, overall linkage disequilibrium on the Z chromosome is similar between species and among regions in all cases except in the region of elevated differentiation in chestnut teal. This region has comparatively high linkage disequilibrium in chestnut teal, but not grey teal. Evidence suggests that linkage disequilibrium tends to be high in speciation with gene flow models within the island of

differentiation (Feder *et al.*, 2012). However, if selection was inhibiting the movement of this region between species, but gene flow was otherwise homogenizing their genomes, then high linkage disequilibrium would be expected in both species. Alternatively, elevated linkage disequilibrium could be a signature of a selective sweep (Kim & Nielsen, 2004). Collectively, the high linkage disequilibrium, low nucleotide diversity, and high fixation of polymorphisms observed in chestnut teal are suggestive of a recent selective sweep (Charlesworth *et al.*, 1993; Nordborg 1997; Charlesworth *et al.*, 1997; Charlesworth, 1998; Cruickshank & Hahn, 2014; Yeaman *et al.*, 2016) within the Z chromosome of this species. Furthermore, low nucleotide diversity rather than elevated absolute divergence is the primary cause of elevated relative divergence between grey and chestnut teal, which is more consistent with post-speciation selection as described by Cruickshank and Hahn (2014). .

Speciation with gene flow

Alternatively to post-speciation selection, absolute divergence should be elevated in islands of differentiation relative to background levels under a speciation-with-gene-flow model (Cruickshank and Hahn, 2014), whereas a selective sweep or purifying selection should not have a significant effect on measures of absolute divergence (Charlesworth, 1998). We found absolute divergence to be approximately 1.2 times higher in the island relative to the rest of the Z chromosome. Although this difference between regions on the Z chromosome is marginally significant ($P = 0.06$), absolute divergence is reduced for the Z chromosome compared to the autosomes (Fig. 5), which is inconsistent with a model of speciation with gene flow.

In addition, under a model of speciation with gene flow there should be no significant advantage to maintaining migration in the island of divergence, whereas including migration should provide a significantly better fit between the model and the data for other regions of the genome (Feder & Nosil 2013; Via, 2012). Our demographic history analyses using ∂a∂i indicate that we cannot reject a strict isolation model for the Z chromosome, suggesting that the amount of migration, if any at all, is insignificant in regards to the power to fit the model to the data. In contrast, the autosomal data suggest that a model of migration fits the data significantly better than a strict isolation model (P= 0.01). This evidence seems consistent with a speciation with gene flow model, in that there is reduced effective gene flow for the Z chromosome relative to genomic levels of genetic exchange. However, it is noteworthy that both the region of elevated divergence and the regions of low divergence on the Z chromosome were consistent with strict isolation, whereas we expected that a migration model would provide a better fit for the regions of low divergence. One possible explanation for these demographic results is that the power for rejecting the simpler model was much lower for the Z chromosome (N = 157 variable loci) than for autosomes (N = 3,101 variable loci). A biological explanation could be that the selective sweep in chestnut teal erased any signatures of past gene flow, which would reduce the overall evidence of gene flow for this region.

CONCLUSION

A great deal of recent work has supported two major models of speciation, the speciation with gene flow, in which divergent selection reduces gene flow locally within the genome, and selection after speciation, in which selection causes a reduction in nucleotide diversity within a species (Via, 2012; Nosil & Feder 2012; Cruickshank and Hahn, 2014). Both of these models result in incidental islands of high relative divergence.

Here we confirmed an earlier finding of elevated Z divergence (Dhami et al., 2016) and provide strong evidence of an island of differentiation on the Z chromosome. Remarkably genetic differentiation between grey and chestnut teal is nearly absent in mitochondrial DNA, autosomal DNA, and the 3'- and 5'- ends of the Z chromosome (Joseph *et al.*, 2009; Dhami *et al.*, 2013; Dhami *et al.*, 2016), whereas the island of differentiation on the q-arm of the Z chromosome harbors fixed differences between species. Although we cannot conclusively reject a model of speciation with gene flow as the cause of this island, the data seem more consistent with an alternative model of post speciation selection acting in the chestnut teal. Thus, these teal seem to be characterized by a recent divergence (Joseph *et al.,* 2009), during which there has been insufficient time for the accumulation of genetic differences via genetic drift, and strong selection acting on the Z chromosome of the chestnut teal. Dhami *et al*. (2016) hypothesized that this divergence might be associated with the transition from a sexually monochromatic ancestor to the sexually dichromatic plumage of the chestnut teal, but additional data are needed to test this hypothesis. Regardless, the ddRAD-seq data unambiguously demonstrate heterogeneity in genomic divergence, as well as heterogeneity in Z divergence, and that selection is playing a role in the genetic divergence of these species, especially on the Z chromosome in chestnut teal. Thus, these teal species are excellent candidates for investigating whether and how selection is playing a role in speciation, which could yield important insights into mechanisms of rapid speciation.

LITERATURE CITED

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* **12**, 1805-1814.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Backström N, Forstmeier W, Schielzeth H, *et al.* (2010) The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* **20**, 485-495.

Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21** 263 – 265.

Bradbury IR, Bowman S, Borza T, *et al.* (2014) Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in atlantic cod. *PLoS One* **9**, e106380.

Chapman MA, Hiscock SJ, Filatov DA (2016) The genomic bases of morphological divergence and reproductive isolation driven by ecological speciation in Senecio (Asteraceae). *Journal of Evolutionary Biology* **29**, 98-113.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538-543.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-1303.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet Res* **70**, 155-174.

Coffman AJ, Hsieh PH, Gravel S, Gutenkunst RN (2016) Computationally efficient composite likelihood statistics for demographic Inference. *Molecular Biology and Evolution* **33**, 591-593.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**, 3133-3157.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One* **9**, e106713.

DeFaveri J, Jonsson PR, Merila J (2013) Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. *Evolution* **67**, 2530-2546.

Delmore KE, Hubner S, Kane NC, *et al.* (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Mol Ecol* **24**, 1873-1888.

Dhami KK, Joseph L, Roshier DA, Heinsohn R, Peters JL (2013) Multilocus phylogeography of Australian teals (*Anas* spp.): a case study of the relationship between vagility and genetic structure. *Journal of Avian Biology* **44**, 169-178.

Dhami KK, Joseph L, Roshier DA, Peters JL (2016) Recent speciation and elevated Z-chromosome differentiation between sexually monochromatic and dichromatic species of Australian teals. *Journal of Avian Biology* **47**, 92-102.

Edgar RC (2004) MUSCLE: multiple seuqnece aligment with high accuracy and high throughput. *Nucleic Acid Research* **32,** 1792 – 1797.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729-1747.

Feldman MW, Franklin I, Thomson GJ (1974) Selection in complex genetic systems. *Genetics* **76**, 135-162.

Foll M, Gaggiotti O (2008) A Genome-Scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics* **180**, 977-993.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* **5**, e1000695.

Haldane JBS (1922) Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics* **12**, 101-109.

Hemmer-Hansen J, Nielsen EE, Therkildsen NO, *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Mol Ecol* **22**, 2653-2667.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226-231.

Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philos Trans R Soc Lond B Biol Sci* **367**, 395-408.

Huang H, Rabosky DL (2015) Sex-linked genomic variation and its relationship to avian plumage dichromatism and sexual selection. *BMC Evol Biol* **15**, 199.

Huang Y, Li Y, Burt DW, *et al.* (2013) The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat Genet* **45**, 776-783.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

Johnson KP, Sorenson MD (1999) Phylogeny and biogeography of dabbling ducks (Genus: *Anas*): a comparison of molecular and morphological evidence. *The Auk* **116**, 792-805.

Johnson NA, Lachance J (2012) The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Ann N Y Acad Sci* **1256**, E1-22.

Joseph L, Adcock GJ, Linde C, *et al.* (2009) A tangled tale of two teal: population history of the grey *Anas gracilis* and chestnut teal *A. castanea* of Australia. *Journal of Avian Biology* **40**, 430-439.

Kim Y, Nielsen R. (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**, 1513-1524

Kraus RH, Kerstens HH, Van Hooft P, *et al.* (2011) Genome wide SNP discovery, analysis and evaluation in mallard (Anas platyrhynchos). *BMC Genomics* **12**, 1-11.

Lavretsky P, Dacosta JM, Hernández-Baños BE, *et al.* (2015) Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology* **24**, 5364-5378.

Li MH, Merila J (2010) Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**, 600-610.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Liu S, Lorenzen ED, Fumagalli M, *et al.* (2014) Populatin genoimcs reveal recent speication and rapid evolutiaionry adaption in polar bears. *Cell* **157**, 785-794.

Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philos Trans R Soc Lond B Biol Sci* **363**, 2971-2986.

Mank JE, Ellegren H (2009) Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution* **63**, 1464-1472.

Mank JE, Nam K, Ellegren H (2010) Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution* **27**, 661-670.

McCoy RC, Garud NR, Kelley JL, Boggs CL, Petrov DA (2014) Genomic inference accurately predicts the timing and severity of a recent bottleneck in a nonmodel insect population. *Mol Ecol* **23**, 136-150.

Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* **28**, 659-669.

Michel AP, Sim S, Powell TH, *et al.* (2010) Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci U S A* **107**, 9724-9729.

Nadeau NJ, Whibley A, Jones RT, *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc Lond B Biol Sci* **367**, 343-353.

Nielsen R, Williamson S, Kim Y, *et al.* (2005) Genomic scans for selective sweeps using SNP data. *Genome Res* **15**, 1566-1575.

Nordborg M (1997) Structured coalescent processes on different time scales. *Genetics* **146**, 1501-1514.

Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philos Trans R Soc Lond B Biol Sci* **367**, 332-342.

Nosil P, Feder JL (2013) Genome evolution and speciation: toward quantitative descriptions of pattern and process. *Evolution* **67**, 2461-2467.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24**, 145-156.

Ohta T, Gillespie JH (1996) Development of neutral and nearly neutral theories. *Theor Popul Biol* **49**, 128-142.

Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. *Philos Trans R Soc Lond B Biol Sci* **365**, 185-205.

Pfeifer B, Wittelsbürger U, Ramos Onsins SE, Lercher MJ (2014) PopGenome: An efficient swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*.

Pryke SR (2010) Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution* **64**, 1301-1310.

Renaut S, Grassa CJ, Yeaman S, *et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun* **4**, 1827.

Saetre G-P, Borge T, Lindroos K, *et al.* (2003) Sex chromosome evolution and speciation in Ficedula flycatchers. *Proceedings of the Royal Society B: Biological Sciences* **270**, 53-59.

Schrider D, Shanku AG, Kern AD (2016) Effects of linked selective sweeps on demographic inference and model selection. *Cold Springs Harbor Labratories, bioRxiv*. DOI: http://dx.doi.org/10.1101/047019

Slatkin M (2008) Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* **9**, 477-485.

Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-35.

Sodeland M, Jorde PE, Lien S, *et al.* (2016) "Islands of dvergence" in the atlantic cod genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution* **8**, 1012-1022.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Supple MA, Papa R, Hines HM, McMillan WO, Counterman BA (2015) Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology* **15**, 1-12.

Turelli M, Orr HA (1995) The dominance theory of Haldane's rule. *Genetics* **140**, 389-402.

Via S (2009) Natural selection in action during speciation. *Proc Natl Acad Sci U S A* **106 Suppl 1**, 9939-9946.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos Trans R Soc Lond B Biol Sci* **367**, 451-460.

Yeaman S, Aeschbacher S, Bürger R (2016) The evolution of genomic islands by increased establishment probability of linked alleles. *Molecular Ecology* **25**, 2542-2558.

FIGURES

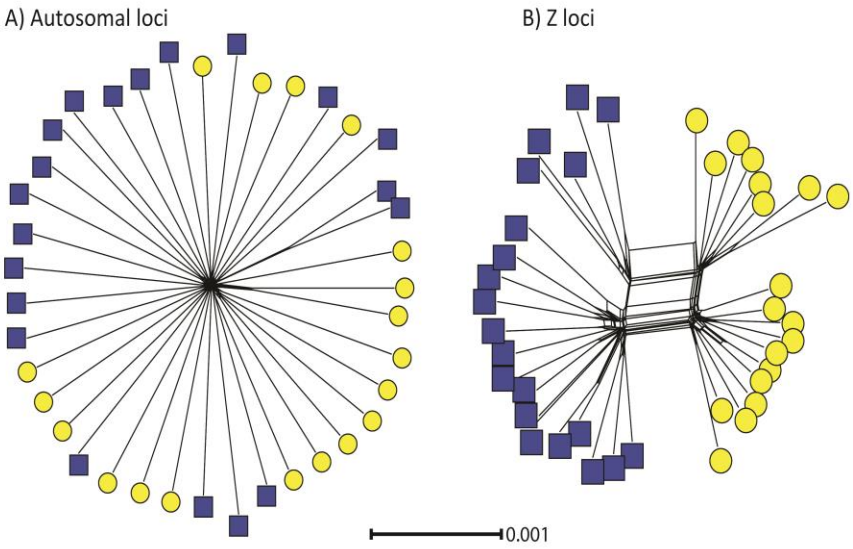

Figure 1. Neighbor-net trees based on 424,812 nucleotides from autosomal loci (1A) and 24,725 nucleotides from the Z-chromosome (1B). Yellow circles represent chestnut teal individuals and blue squares represent grey teal individuals. Note the strong separation between species for the Z-chromosome (1B).

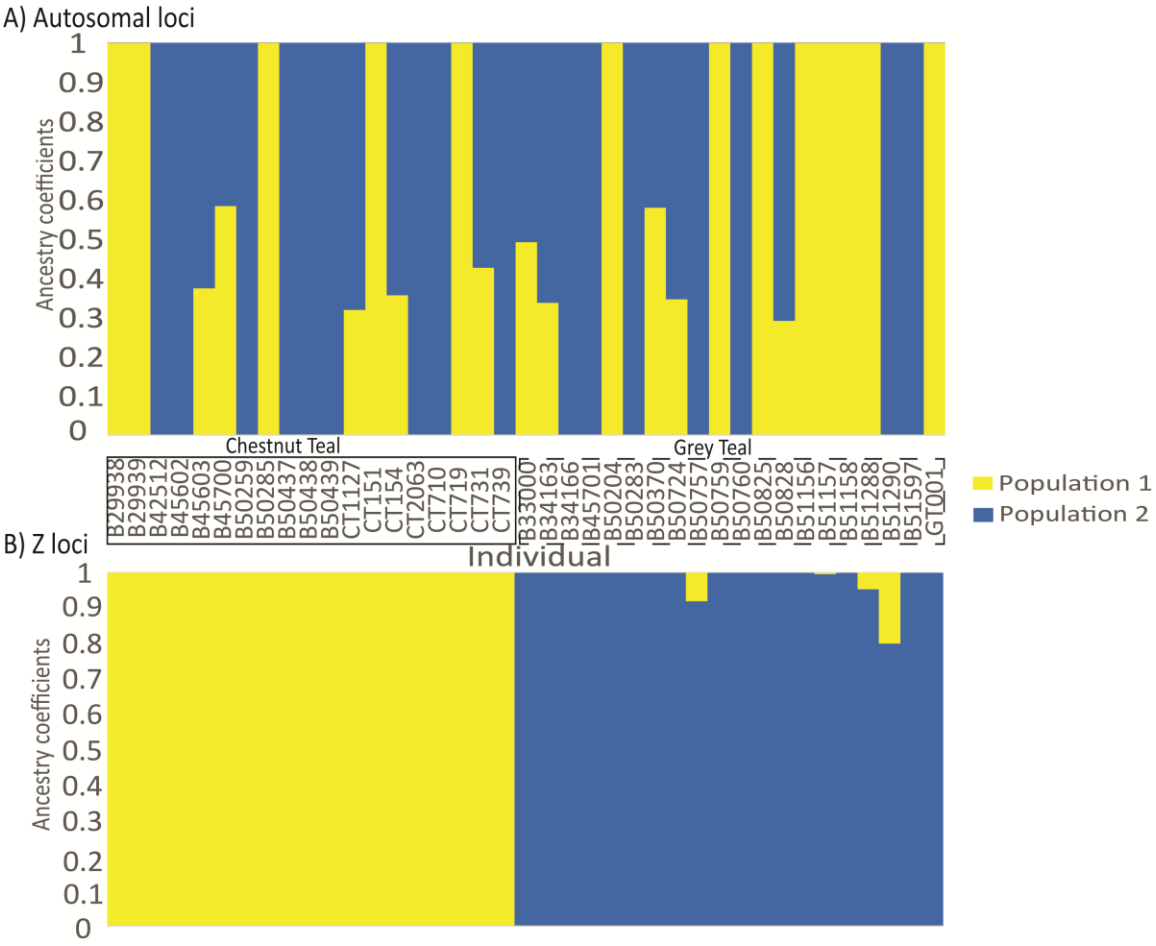

Figure 2. Ancestry coefficients for chestnut and grey teal based on autosomal loci (A) and Z-chromosome loci (B). The yellow bars represents the confidence of that individual being assigned to population 1 which we designate as chestnut teal. Blue bars represent confidence of the individual being assigned to population 2 which we classified as grey teal.

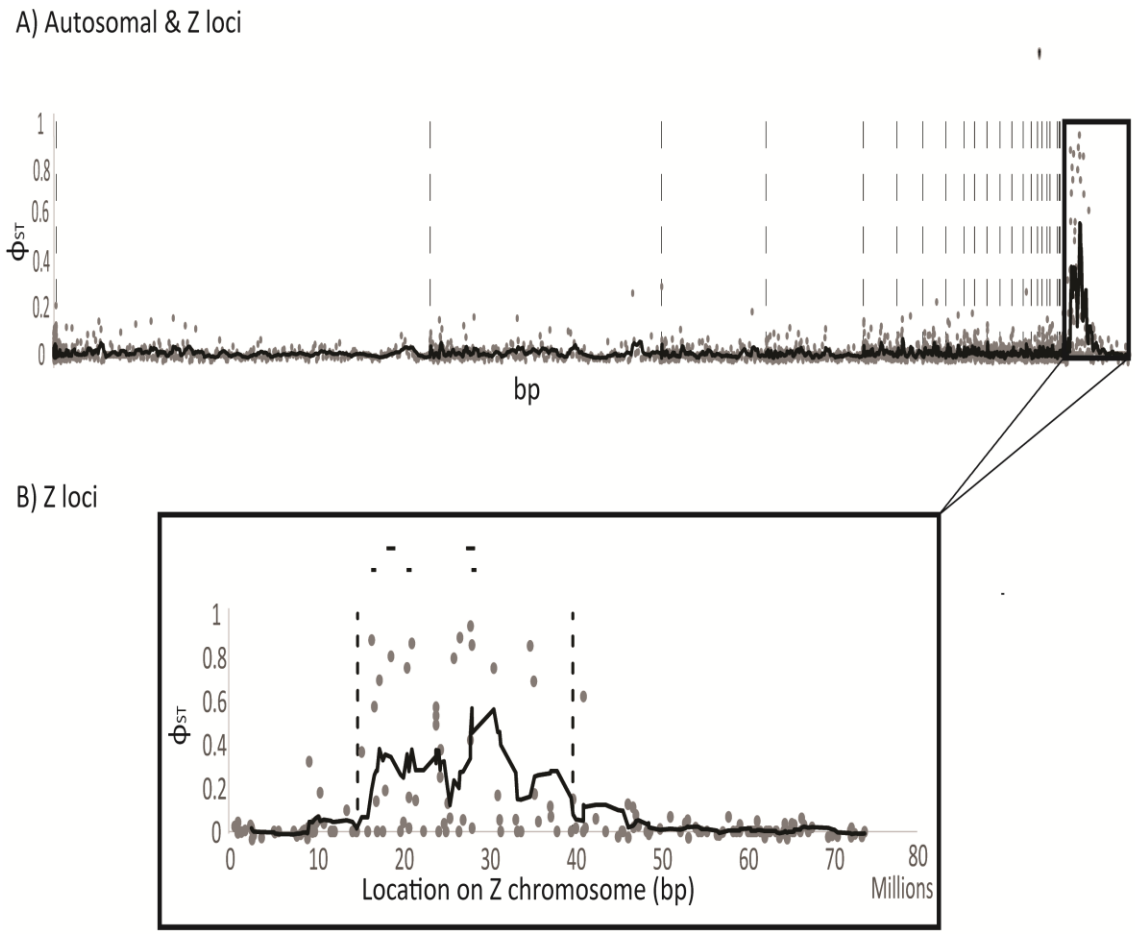

Figure 3. Relative divergence (ΦST) between grey teal and chestnut teal across the autosomes with the Z chromosome on the tail (A) and the Z chromosome magnified (B). Vertical dashed lines indicate chromosomal boundaries in the autosomes (A). Bold black lines across the x-axis represents a moving average of seven loci. Upper horizontal dashed lines above graph denote significant outlier loci relative to background levels of differentiation found using BayeScan (A-B). The lower horizontal dash represents loci of trending elevation but not quite significant whereas (B). NOTE: The dotted black line in 3B denotes the arbitrary boundary of the region of elevated differentiation.

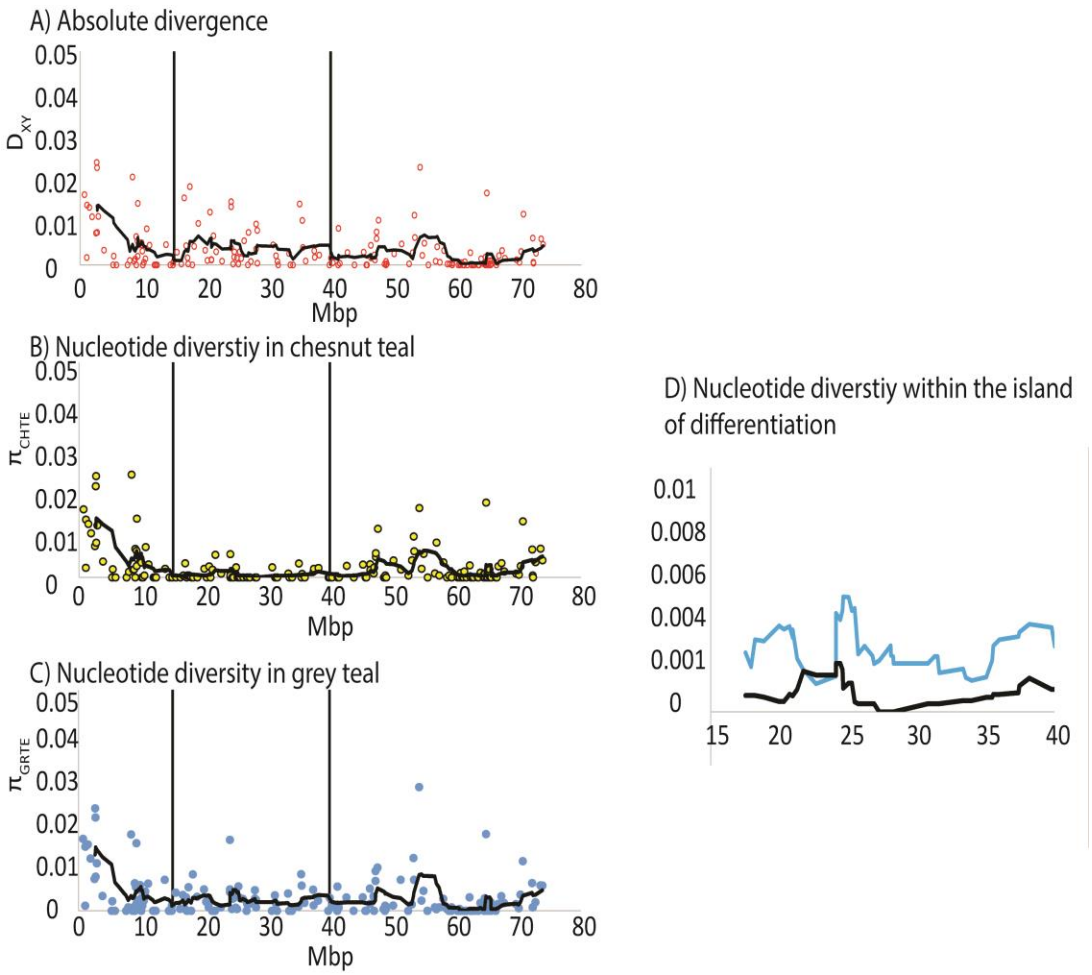

Figure 4. Absolute divergence (dxy) between grey and chestnut teal (A), nucleotide diversity within chestnut teal (B), and nucleotide diversity within grey teal (C) for Z-linked loci. Horizontal lines (A-C) represents moving average of seven loci. Note the significant reduction in nucleotide diversity in the island relative to outside the island for chestnut teal. (D) Comparison of nucleotide diversity between chestnut teal (black line) and grey teal (blue line) in the region of elevated differentiation (represents the moving average of five loci in chestnut and grey teal, respectively). The black vertical line (A-C) represents the same arbitrary region of elevated differentiation in figure 3B.

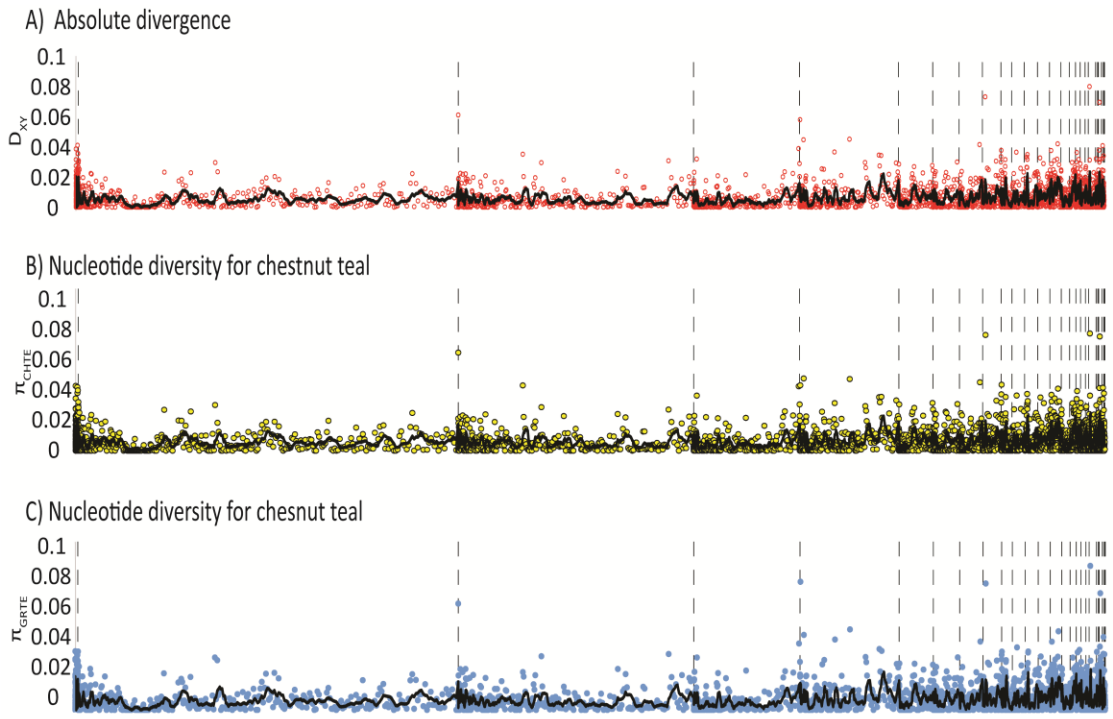

Figure 5. Absolute divergence (dxy; A), nucleotide diversity within chestnut teal (πCHTE; B), and nucleotide diversity within grey teal (πGRTE; C) for autosomal loci. The vertical dotted lines denote chromosomal boundaries, and the bold horizontal line represents the moving average for every seven loci.

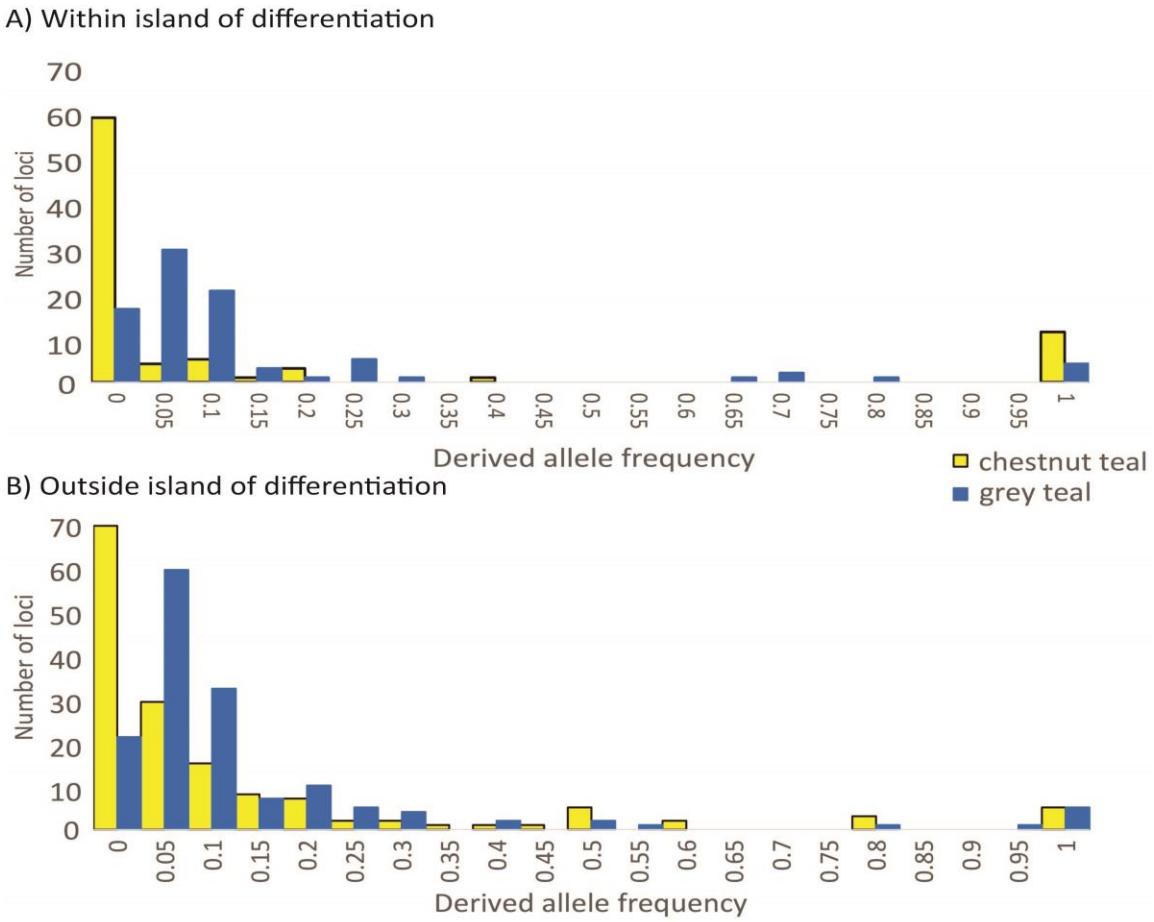

Figure 6. Frequency distribution of the derived SNP in the region of elevated differentiation (A) and outside the region of elevated differentiation (B) for chestnut teal (yellow bars) and grey teals (blue bars).

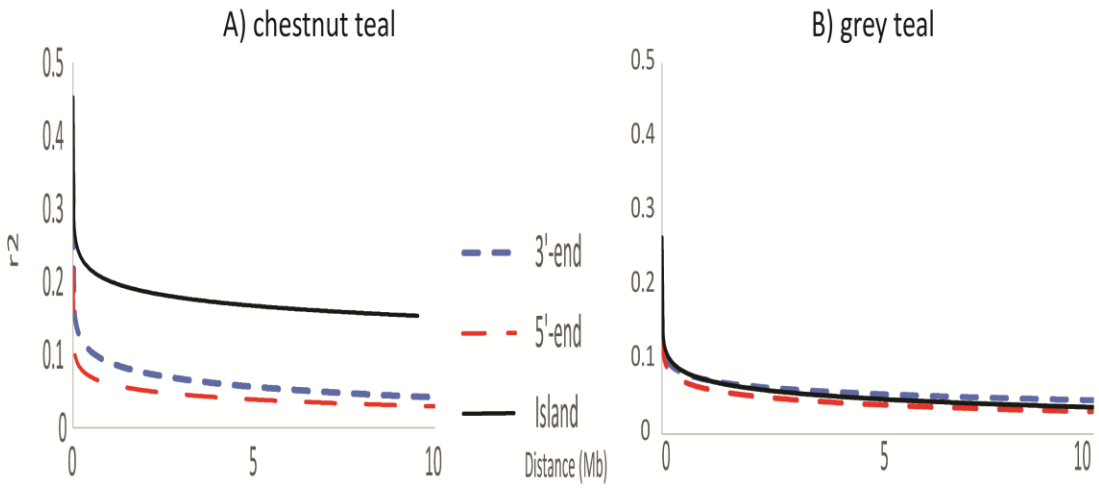

Figure 7. Linkage decay within three regions of the Z chromosome for chestnut teal (A) and grey teal (B). Black dotted line represents a logarithmic regression of linkage (r2) in the island region on the Z chromosome, and red and blue dotted lines represents logarithmic regression of linkage in the 5' and 3' ends, respectively. Note the elevated linkage within the island of differentiation for chestnut teal.