

2013

Genomic analysis of macro- and micro-evolution in the reptilia

<https://hdl.handle.net/2144/15658>

Downloaded from DSpace Repository, DSpace Institution's institutional repository

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS & SCIENCES

Dissertation

**GENOMIC ANALYSIS OF MACRO- AND MICRO-EVOLUTION
IN THE REPTILIA**

by

NICHOLAS GEOFFREY CRAWFORD

B.S., Union College 2001
M.S., San Diego State University 2007

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2013

Approved by

First Reader

Christopher J. Schneider, Ph.D.
Associate Professor of Biology

Second Reader

Sean P. Mullen, Ph.D.
Assistant Professor of Biology

Acknowledgments

As I think it is for most graduate students, writing and researching my dissertation has been one of the most difficult things I've ever done. But, when I have felt the most discouraged, exhausted, and frustrated, I've tried to remember my sister before me. It was her encouragement that convinced me to follow this path and it has been her memory, particularly that of her ferocious tenacity, which has helped me make it to the end. Corinne, thank you for your leadership along the way.

**GENOMIC ANALYSIS OF MACRO- AND MICRO-EVOLUTION
IN THE REPTILIA**

(Order No.)

NICHOLAS G. CRAWFORD

Boston University Graduate School of Arts and Sciences, 2013

Major Professor: Christopher J. Schneider, Associate Professor of Biology

ABSTRACT

Recent advances in high-throughput, genomic sequencing allow unprecedented insight into the evolution of biodiversity. Chapter 1 of this thesis is a phylogenetic study of 1,145 sequenced loci, isolated using a novel high-throughput sequence capture methodology to address the phylogenetic position of turtles within tetrapods. The results reported here unambiguously place turtles as sister to archosaurs and resolve this long-standing question.

Chapter 2 investigates the genetic basis of colorful pigmentation in the Green anole (*Anolis carolinensis*) by sequencing complete transcriptomes from the green dorsal, white ventral and pink dewlap skin. Anoles comprise an adaptive radiation of more than 400 species and color plays a central role in their ecology and evolution, but little is known about the genetic basis of colorful pigmentation in any vertebrate. This study identified 1,719 differentially expressed genes among the three differently colored tissues. Twenty-three of these genes are involved in melanin, pteridine, and carotenoid pigmentation pathways that contribute to the coloration of anole skin. Identifying

candidate genes for colorful pigmentation is a significant advance that opens the field for comparative analysis in other taxa.

To determine if the genes identified in Chapter 2 are involved in population divergence and speciation, Chapter 3 investigates the complete genomes of twenty individuals from two closely related subspecies of *Anolis marmoratus*. While the two subspecies differ markedly in pigmentation, this study found few genetic differences between populations except in five regions of the genome, which together contained 447 genes. Of these genes, only two, melanophilin (*mlph*) and ‘cluster of differentiation 36’ (*cd36*), are associated with pigmentation. The intersection of the genes identified in Chapter 2 and Chapter 3 includes both *cd36* and *mlph*, suggesting that both are involved in divergence of coloration. *Cd36* is of particular interest because it regulates the uptake of carotenoid pigments and is an important candidate gene contributing to carotenoid pigmentation.

Together, this research demonstrates the power of genomic approaches to address fundamental questions in systematics, micro-evolution, and speciation. The findings bolster the emerging field of phylogenomics and broadly impact future research into the genetic basis of coloration in vertebrates.

Table of Contents

Acknowledgments	iii
Abstract	iv
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
Chapter 1: More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs	1
Summary	1
<i>Introduction</i>	1
<i>Methods</i>	3
<i>Results</i>	4
<i>Discussion</i>	5
<i>Figures and Tables</i>	8
Chapter 2: Genetics of Colorful Pigmentation in the Green Anole (<i>Anolis carolinensis</i>).	12
<i>Summary</i>	12
<i>Introduction</i>	13
<i>Results</i>	16
Gene Ontology terms.	17
Melanin associated genes.	18
Pteridine associated genes.	20
Carotenoid associated genes.	21

Genes associated with blue structural colors.	22
qPCR of pteridine genes.	23
<i>Discussion</i>	24
<i>Methods</i>	28
Phenotypic measurements.	28
Pooled RNA-sequencing.	29
Single Sample RNA sequencing.	30
Quantitative PCR.	30
<i>Figures and Tables</i>	32
<i>Supplemental figures, tables, and methods</i>	40
Chapter 3: Genomics of Coloration in <i>Anolis marmoratus</i>	58
<i>Summary</i>	58
<i>Introduction</i>	59
<i>Methods</i>	63
Sample Collection.	63
Library Construction and Analysis.	63
Population Genomic Analysis.	65
<i>Results</i>	66
Sequencing and Structural variation.	66
Analysis of Outliers.	67
Genomic Architecture.	68
<i>Discussion</i>	71
Conclusion.	73

<i>Figures and Tables</i>	76
<i>Supplemental figures, tables, and methods</i>	83
Appendix	88
<i>SMOGD: Software for the Measurement of Genetic Diversity</i>	89
Summary	89
Main Text	89
<i>Pypgen: Calculating a Diversity of Fixation Indexes Across Genomes</i>	91
Summary	91
Introduction	91
Pypgen	93
Bibliography	95
Curriculum Vitae	110

List of Tables

Table 1-1	Summary information including sample ID and quantity of sequence data used in phylogenetic analysis of ultra-conserved elements of reptiles.	11
Table 2-1	Counts of genes that are significantly down or up regulated in the dewlap, dorsal, and ventral skin in the RNA-seq analyses .	35
Table 2-2	Significantly differentially expressed pigmentation genes identified from mRNA sequence profiles of dewlap, dorsal and ventral skin of <i>Anolis carolinensis</i> .	37
Table 2-3	Analysis of covariance of the pteridine pathway in qPCR of dewlap, dorsal, and ventral skin of <i>Anolis carolinensis</i> .	38
Table 2-4	Multivariate analysis of covariance of pteridine genes <i>pts</i> , <i>spr</i> , and <i>pam</i> in qPCR in <i>Anolis carolinensis</i> .	39
Supp. Table 2-1	Significant GO categories in expressed gene sets	44
Supp. Table 2-2	Candidate coloration genes identified in this study	50
Supp. Table 2-3	Sequences of quantitative PCR primers used in this study	51
Supp. Table 3-1a	The 19 fixed SNVs with at least 5 samples per population within coding sequences	83
Supp. Table 3-1b	Table 3-1a continued: Ensembl Gene IDs to Gene Names	84
Supp. Table 3-2	Genes within genomic islands of divergence containing at least one fixed SNV	87

List of Figures

Figure 1-1	Three hypotheses of turtle relationship to other reptiles .	9
Figure 1-2	Reptile phylogeny estimated from 1,145 ultra-conserved loci (UCEs).	10
Figure 2-1	The phenotype of the green anole (<i>Anolis carolinensis</i>) showing different colored skin of the dorsum, venter and dewlap.	32
Figure 2-2	Box plots of gene expression for genes in the pteridine pigment pathway, <i>pts</i> , <i>spr</i> , and <i>pam</i> , analyzed by qPCR from male <i>A. carolinensis</i> ranging in size from juveniles – adults.	33
Figure 2-3	Scatter plots of gene expression for genes in the pteridine pigment pathway from male <i>A. carolinensis</i> ranging in size from juveniles – adults	34
Supp. Figure 2-1	Dewlap vs dorsal scatter plot of FPKM values	52
Supp. Figure 2-2	Dewlap vs ventral scatter plot of FPKM values	53
Figure 3-1	Images and range map of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> showing the distribution of phenotypic variation among populations.	76
Figure 3-2	Box plots showing mean and variance of G''_{ST} and Tajima's D for each chromosome derived from analysis of whole genome sequences from populations of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> .	77

Figure 3-3	Histogram of G''_{ST} values for 5kb windows across the complete genomes of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> .	78
Figure 3-4	Density plot of Tajima's D for 5kb windows across the complete genomes of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> .	79
Figure 3-5	Circos plot of summary statistics including G''_{ST} , Tajima's D , Linkage Disequilibrium, and fixed SNVs across the complete genomes of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> . Gene structure for specific regions is also displayed.	81
Figure 3-6	Bar plot of mean G''_{ST} , Tajima's D and Linkage Disequilibrium within and outside of islands of divergence identified from comparison of the complete genomes of <i>Anolis marmoratus marmoratus</i> and <i>Anolis marmoratus speciosus</i> .	82

List of Abbreviations

α	alpha
β	beta
$\Delta\Delta Ct$	Change of change in cycle threshold
μg	Microgram
μL	Microliter
ABYSS	"Assembly By Short Sequences" - a de novo, parallel, paired-end sequence assembler
aim1	Protein absent in melanoma 1
AMNH	American Museum of Natural History
ANCOVA	Analysis of covariance
ANNOVAR	Functional annotation of genetic variants from high-throughput sequencing data (computer software)
anoCar2	Anolis carolinensis genome version 2
<i>aox1</i>	Aldehyde oxidase 1
<i>aox3</i>	Aldehyde oxidase 3
AWTY	"Are We There Yet?" - A system for graphical exploration of Markov chain Monte Carlo (MCMC) convergence in Bayesian phylogenetic inference
BED	Browser Extensible Data format
bp	Base-pair
C	Celsius
<i>cd36</i>	Cluster of differentiation 36
cDNA	Complementary DNA (the product of reverse transcription)

	of RNA to DNA)
cm	Centimeter
<i>csflr</i>	Colony Stimulating Factor Receptor 1
Ct	Cycle threshold
<i>dct</i>	Dopachrome tautomerase
D_{est}	Estimator of actual differentiation (Jost 2008)
DNA	Deoxyribonucleic acid
DNA Pol 1	Deoxyribonucleic acid polymerase 1
dNTP	Deoxyribonucleotide
dsDNA	Double stranded deoxyribonucleic acid
DTT	Dithiothreitol
e-value	Expect value
EB	Elution buffer
Ensembl	Genomics resource, the product of a joint scientific project between the European Bioinformatics Institute and the Wellcome Trust Sanger Institute
ESS	Effective Sample Size
EtOH	Ethanol
FDR	False Discovery Rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
F_{ST}	Fixation Index (Wright 1949)
g	g-force
GATK	Genome Analysis Tool Kit

GO	Gene Ontology
<i>gsta2</i>	Glutathione S-transferase A2
GTP	guanosine-5'-triphosphate
G_{ST}	Relative differentiation (Masatoshi Nei 1973)
G'_{ST}	Standardized of genetic differentiation (Hedrick 2005)
G''_{ST}	Corrected G'_{ST} (Meirmans & Hedrick 2011)
H ₂ O	H ₂ O: water
HPLC	High pressure liquid chromatography
H _S	Within-population gene diversity(Masatoshi Nei 1973)
H _{S est}	Nearly unbiased estimator of within-subpopulation heterozygosity (Nei & Chesser 1983)
H _T	Total gene diversity(Masatoshi Nei 1973)
H _{T est}	Nearly unbiased estimator of total-subpopulation heterozygosity (Nei & Chesser 1983)
Illumina GAIIx	Illumina GAIIx: Illumina Genome AnalyzerIIx sequencer
<i>k</i>	The number of sampled populations
K	"An overlap between the last and the first $k-1$ characters of two adjacent [nucleotide sequences]" (Simpson <i>et al.</i> 2009)
KAPA	Kapa Biosystems (A biotech company)
kbp	Kilobase-pairs
Klenow 3'-5'exo	N-terminal truncation of DNA Polymerase I which retains polymerase activity, but has lost the 5'→re' exonuclease activity
<i>krt80</i>	Keratin 80

LG	Linkage Group
LOESS	Locally weighted scatterplot smoothing
MANCOVA	Multivariate analysis of covariance
max	Maximum
mg	Milligrams
MgCl ₂	Magnesium chloride
microRNA	A small non coding RNA molecule
min	Minute
miRNA	A small non coding RNA molecule (e.g., microRNA)
<i>mlph</i>	Melanophilin
mm	Millimeter
MMix	Master Mix
MrAIC	Perl script for calculating AIC, AICc, BIC, and Akaike weights for sequence alignments
MrBayes	Software program which performs Bayesian inference of phylogeny.
MUSCLE	MULTiple Sequence Comparison by Log-Expectation
n	Number of populations
\tilde{N}	Harmonic mean of population sizes
N	Total sample size
NaCl	Sodium chloride
NaOAc	Sodium acetate
NEB	New England Biolabs

nm	Nanometers
NS	Nonsynonymous
NSF	National Science Foundation
<i>oca2</i>	Oculocutaneous albinism II
oligo-dT	An oligonucleotide made of deoxy-thymine nucleotides
p-value	Probability value
pam	Protein Associated with Myc
PCR	Polymerase Chain Reaction
<i>ped4</i>	Phosphodiesterase 4
PEG	Polyethylene glycol
<i>plin2:</i>	Perilipin 2
<i>pmel</i>	Silver gene
<i>pnp4a</i>	Purine nucleoside phosphorylase 4a
<i>pts</i>	6-pyruvoyltetrahydropterin synthase
PX-2	Pulsed Xenon version 2 (light source)
qPCR	Quantative PCR
RAD	Restriction Site Associated DNA markers
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Room temperature
RT	Reverse transcriptase
rxn	Reaction

S	Synonymous
SD	Standard deviation
SG	Stop or gain of function
<i>silv</i>	Silver gene
<i>slc24a5</i>	Solute carrier family 24 member 5
SMOGD	Software for the measurement of genetic diversity
SNV	Single nucleotide variant
<i>spr</i>	Sepiapterin reductase
SPRI	SPRI beads - biotinylated beads
STAR	Species tree estimation using average ranks of coalescences
STEAC	Species tree estimation using average coalescence times
SVL	Snout-vent length
tblastx	Tools for searching translated nucleotide databases using a translated nucleotide query
TE	Tris-Ethylenediaminetetraacetic acid
Tracer	A program for analysing the trace files generated by Bayesian MCMC
Tris-HCl	Tris(hydroxymethyl)aminomethane-hydrochloride
TRIzol	A chemical used in RNA/DNA/protein extraction
<i>tryp1</i>	Tyrosinase-related protein 1
<i>txn15</i>	Thioreductase-like 5 or clot
<i>tyr</i>	Tyrosinase
UCE	Ultra conserved element

UCSC	University of California, Santa Cruz
USB	Universal Serial Bus
UV-VIS	Ultra Violet - Visual Spectrum
vs	versus

Chapter 1: More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs

Summary

We present the first genomic-scale analysis addressing the phylogenetic position of turtles, using over 1,000 loci from representatives of all major reptile lineages including tuatara. Previously, studies of morphological traits positioned turtles either at the base of the reptile tree or with lizards, snakes, and tuatara (lepidosaurs), whereas molecular analyses typically allied turtles with crocodiles and birds (archosaurs). A recent analysis of shared microRNA families found that turtles are more closely related to lepidosaurs. To test this hypothesis with data from many single-copy nuclear loci dispersed throughout the genome, we used sequence capture, high-throughput sequencing, and published genomes to obtain sequences from 1,145 ultraconserved elements (UCEs) and their variable flanking DNA. The resulting phylogeny provides overwhelming support for the hypothesis that turtles evolved from a common ancestor with birds and crocodylians, rejecting the hypothesized relationship between turtles and lepidosaurs.

Introduction

The evolutionary origin of turtles has confounded the understanding of vertebrate evolution (Lee *et al.* 2004) (see **Figure 1-1**). Historically, turtles were thought to be

early-diverging reptiles, called anapsids, based on their skull morphology and traits such as dermal armor (Lee 1997). Recent morphological studies that included soft tissue and developmental characters (Rieppel 1999) allied turtles with lepidosaurs, a group including squamates (lizards and snakes) and tuataras. However, homoplasy stemming from the derived skeletal specializations of turtles limits the utility of phylogenetic inference based on morphological data to resolve turtle placement (Hedges & Poling 1999; Janke *et al.* 2001).

Molecular studies using mitochondrial (Hedges *et al.* 1990; Zardoya & Meyer 1998; Kumazawa & Nishida 1999; Janke *et al.* 2001; Rest *et al.* 2003) and nuclear DNA (Hedges & Poling 1999; Cao *et al.* 2000; Iwabe *et al.* 2005; Hugall *et al.* 2007; Shedlock *et al.* 2007; Katsu *et al.* 2009; Tzika *et al.* 2011; Shen *et al.* 2011) typically place turtles sister to archosaurs (crocodilians and birds) (Figure 1-1). This molecular hypothesis was recently contradicted by a phylogeny reconstructed from microRNAs (Lyson *et al.* 2012) that allied turtles with lepidosaurs. Lyson *et al.* (2012) suggested that prior molecular evidence for a turtle-archosaur relationship may be the result of analytical artifacts. If true, the hypothetical relationship between turtles and lepidosaurs (Ankylopoda) should appear throughout the genomes of these organisms.

Here we test the Ankylopoda hypothesis and address the evolutionary origin of turtles. We reconstruct a reptile phylogeny using ultraconserved elements (Bejerano *et al.* 2004) and their flanking sequence (hereafter UCEs) that we obtained using sequence capture of DNA from a tuatara and two species each of crocodilians, squamates, and turtles (**Table 1-1** We used UCEs because they are easily aligned portions of extremely

divergent genomes (Siepel *et al.* 2005), allowing many loci to be interrogated across evolutionary timescales, and because sequence variability within UCEs increases with distance from the core of the targeted UCE (Faircloth *et al.* 2012), suggesting that phylogenetically informative content in flanking regions can inform hypotheses spanning different evolutionary timescales. To break up long branches and mitigate potential problems with long-branch attraction, we selected species representing the span of diversity within major reptilian lineages (i.e., two of the most divergent crocodylians, lepidosaurs including tuatara, and turtles).

Methods

We enriched DNA libraries prepared with Nextera kits (Epicentre, Inc.) using a synthesis (Mycroarray, Inc. or Agilent, Inc.) of RNA probes (Faircloth *et al.* 2012) targeting 2,386 ultraconserved elements and their flanking sequences. We generated sequences for each enriched library using single-end, 100-base sequencing on an Illumina GAIIx. After quality filtering we assembled reads into contigs using Velvet (Zerbino & Birney 2008), and we matched contigs to the UCE loci, removing duplicate hits to avoid paralogous loci. We generated alignments using MUSCLE (Edgar 2004), and we excluded loci having missing data in any taxon. Following alignment we estimated the appropriate finite-sites substitution model for each locus using MrAIC.

We prepared a concatenated dataset by partitioning loci by substitution model prior to analysis using two runs of MrBayes (Ronquist *et al.* 2012) for 5,000,000 iterations (4 chains per run; burn-in: 50%; thinning: 100). We also used each alignment to estimate gene trees incorporating 1,000 two-stage bootstrap replicates of both alignments

and characters (Seo *et al.* 2005), which we integrated into STEAC and STAR (Liu *et al.* 2009) species trees. Additional details concerning UCE sequence capture methods and phylogenetic methods are available in Faircloth *et al.* (2012).

Results

We enriched genomic DNA for UCEs in corn snake (*Pantherophis guttata*), African helmeted turtle (*Pelomedusa subrufa*), painted turtle (*Chrysemys picta*), American alligator (*Alligator mississippiensis*), saltwater crocodile (*Crocodylus porosus*), and tuatara (*Sphenodon tuatara*) (Table 1-1). We sequenced a mean of 4.9 million reads from each library, and from these reads we assembled an average of 2,648 (\pm 314 SD) contigs. We supplemented these taxa with UCEs extracted from the chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), Carolina anole lizard (*Anolis carolinensis*), and human (*Homo sapiens*) genome sequences. We combined the *in silico* and *in vitro* data and generated alignments across all taxa and excluded all loci having missing data from any taxon. This resulted in 1,145 individual alignments with a mean length of 406 bp (\pm 100 bp SD) per alignment, totaling 465 Kbp of sequence. Tracer showed that both Bayesian analyses converged quickly, having ESS scores for log likelihood of 170 and 220. AWTY (<http://ceb.csit.fsu.edu/awty>) showed zero variance in the tree topology throughout either run. Bayesian analysis of concatenated alignments and species-tree analysis of 1,145 independent gene histories showed turtles to be the sister lineage of extant archosaurs with complete support (**Figure 1-2**). Removing the snake, which had a very long branch, and re-running all analyses did not change the results.

Discussion

Genomic-scale phylogenetic analysis of 1,145 nuclear UCE loci agreed with most other molecular studies (Zardoya & Meyer 1998; Hedges & Poling 1999; Kumazawa & Nishida 1999; Janke *et al.* 2001; Rest *et al.* 2003; Iwabe *et al.* 2005; Hugall *et al.* 2007; Shedlock *et al.* 2007; Katsu *et al.* 2009; Tzika *et al.* 2011; Shen *et al.* 2011), supporting a sister relationship between turtles and archosaurs. We found no support for the turtles/lizard relationship predicted by the Ankylopoda hypothesis (Lyson *et al.* 2012) (Figure 1-2). The combination of taxonomic sampling, the genome-wide scale of the sampling, and the robust results obtained, regardless of analytical method, indicates that the turtle-archosaur relationship is unlikely to be caused by long-branch attraction or other analytical artifacts.

Although our results corroborate earlier studies, many of these studies did not include tuatara. Because tuatara is an early-diverging lepidosaur, it is important to include this taxon in studies of turtle evolution as it breaks up the long-branch leading to squamates (Figure. 1-2b). Of the studies including tuatara, two (Rest *et al.* 2003; Hugall *et al.* 2007) found results similar to this study, but both were based on a single locus. The third study (Hedges & Poling 1999) was unable to produce a well-resolved tree from four nuclear genes when the authors included tuatara in the dataset. Our study is the first to produce a well-resolved reptile tree that includes the tuatara and multiple loci.

The discrepancy between our results showing a strong turtle-archosaur relationship and microRNA (miRNA) results, which showed a strong turtle-lepidosaur relationship, may be due to several factors. (Lyson *et al.* 2012) used the presence of four

miRNA gene families, detected among turtles and lepidosaurs and undetected in the other taxa analyzed, to support the turtle-lepidosaur relationship. Because complete genomes are unavailable for turtles, tuatara, and crocodylians, and because expressed miRNA data are lacking for most reptiles, the authors collected miRNA sequences from small RNA expression libraries. miRNAs have tissue and developmental-stage specific expression profiles (Lagos-Quintana *et al.* 2002; Landgraf *et al.* 2007), which could make the detection of certain miRNAs challenging. Because preparing and sequencing libraries is a biased sampling process, the detection probability for specific targets is variable, and some miRNAs are likely to be more easily detected than others. Thus, failures to detect miRNA families are not equivalent to the absence of miRNA families (MacKenzie *et al.* 2002). We suggest that at least some of the four miRNA families currently thought to be unique to lizards and turtles may be present but as yet undiscovered in other reptiles.

This work is the first to investigate the placement of turtles within reptiles using a genomic-scale analysis of single-copy DNA sequences and a complete sampling of the major relevant evolutionary lineages. Because UCEs are conserved across most vertebrate groups (Faircloth *et al.* 2012) and found in groups including yeast and insects (Siepel *et al.* 2005), our framework is generalizable beyond this study and relevant to resolving ancient phylogenetic enigmas throughout the tree of life (McCormack *et al.* 2011; Faircloth *et al.* 2012). This approach to high-throughput phylogenomics – based on thousands of loci – is likely to fundamentally change the way that systematists gather and analyze data.

Additional Information

We provide all data and links to software via Dryad repository

(doi:10.5061/dryad.75nv22qj) and GenBank (JQ868813 - JQ885411).

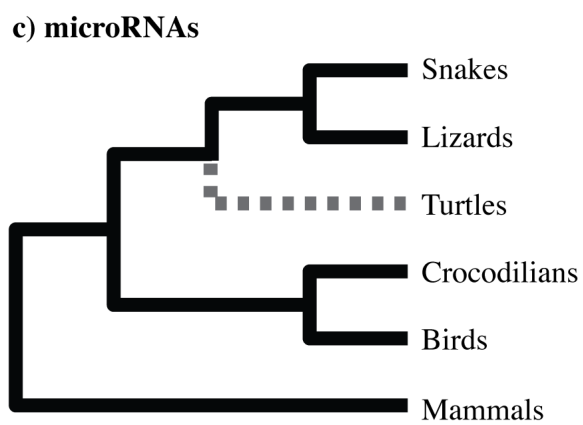
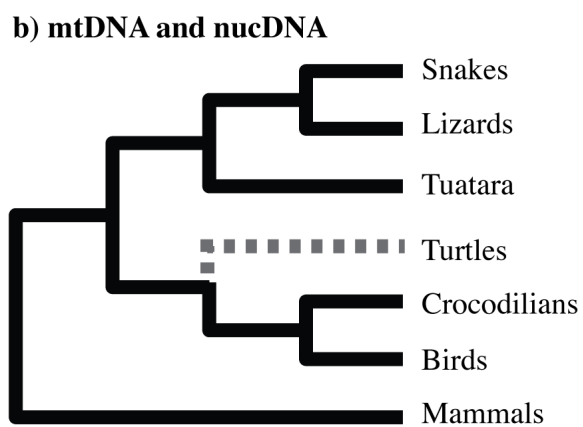
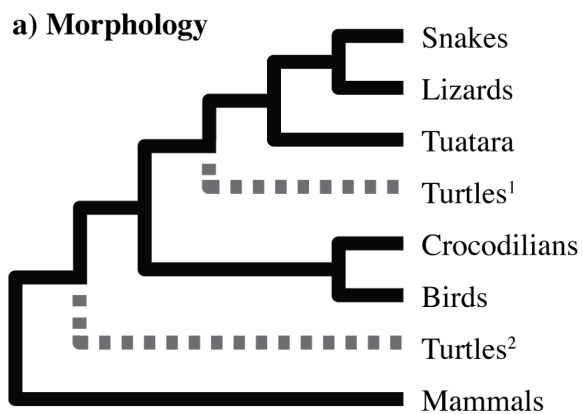
Figures and Tables

Figure 1-1 Three hypotheses for the placement of turtles in reptile phylogeny. a) Depicts the hypotheses based on morphological characters: turtles as the sister group of all remaining reptiles (Lee 1997)² or turtles as the sister group to lepidosaurs (Rieppel 1999)¹. b) Depicts the hypothesis of a turtle-archosaur alliance based on sequences of mitochondrial and nuclear genes (Zardoya & Meyer 1998; Hedges & Poling 1999; Kumazawa & Nishida 1999; Janke *et al.* 2001; Rest *et al.* 2003; Iwabe *et al.* 2005; Hugall *et al.* 2007; Shedlock *et al.* 2007; Katsu *et al.* 2009; Tzika *et al.* 2011; Shen *et al.* 2011). c) Depicts the tree derived from analysis of the presence or absence of specific miRNA loci (Lyson *et al.* 2012).

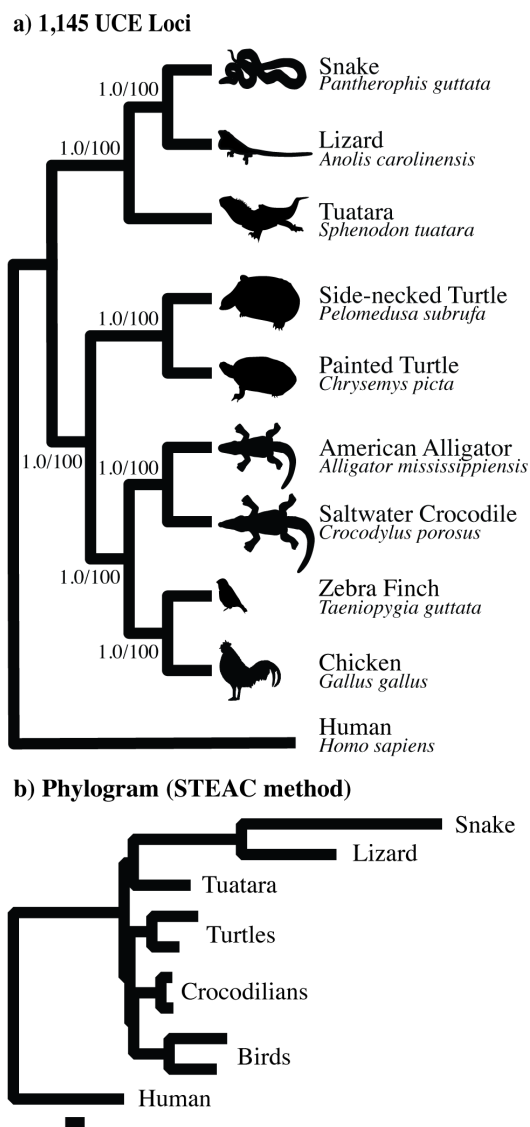


Figure 1-2 a) Reptilian phylogeny estimated from 1,145 ultra-conserved loci using Bayesian analysis of concatenated data and species-tree methods, yielding identical topologies. Node labels indicate posterior probability/bootstrap support. b) Phylogram of the UCE phylogeny generated with STEAC. Branch lengths represent average coalescence units.

Common Name	Binomial	Specimen ID /Genome Build	Reads	Assembled UCEs
African helmeted turtle	<i>Pelomedusa subrufa</i>	H20145 ^a	11,200,032	1972
American alligator	<i>Alligator mississippiensis</i>	HCD-2620 ^a	3,528,983	2320
Carolina anole	<i>Anolis carolinensis</i>	H16061 ^a	3,100,147	2111 ^d
Corn snake	<i>Pantherophis guttata</i>	H15909 ^a	3,362,738	2168
Human	<i>Homo sapiens</i>	UCSC hg19	NA	1748
Painted turtle	<i>Chrysemys picta</i>	H2662 ^a	4,467,644	2261
Red junglefowl	<i>Gallus gallus</i>	UCSC galGal3	NA	2360 ^d
Saltwater crocodile	<i>Crocodylus porosus</i>	LM-67 ^b	3,261,088	2218
Tuatara	<i>Sphenodon tuatara</i>	UMFS-10956 ^c	5,651,932	2199
Zebra finch	<i>Taeniopygia guttata</i>	UCSC taeGut1	NA	2345 ^d

^afrom the LSU museum of Natural Science; ^bfrom the Darwin Crocodile Farm courtesy of L. Miles, S. Isberg, and C. Moran; ^cfrom the University of Michigan Museum of Zoology courtesy of R. Nussbaum and G. Schneider.

^dAlthough we identified 2,386 UCEs in these organisms, from which we designed capture probes, due to slight adjustments to matching and filtering algorithms we only recover ca. 98% of these UCEs when re-screening these genomic sequences.

Table 1-1 UCSC genome build or specimen ID for each sample, the number of ~100 bp sequence reads, and the total number of UCEs assembled.

Chapter 2: Genetics of Colorful Pigmentation in the Green Anole (*Anolis carolinensis*).

Summary

Coloration plays a central role in local adaptation and speciation in many organisms and likely evolves by both ecological and sexual selection. Dozens of studies have focused on the genetic basis of cryptic coloration and its importance in ecological selection. However, much less is known about the genetic basis of conspicuous coloration, which is often under sexual selection. Here we investigate the genetic basis of coloration of both cryptic and conspicuous coloration in *Anolis* lizards. Anoles use cryptic dorsal coloration to hide from predators but also possess a dewlap, a conspicuously colored throat fan, which plays an important role in social signaling and species recognition. However, almost nothing is known about the genetic basis of coloration in anoles. In this study we sequenced the skin transcriptomes of the dewlap, dorsal, and ventral skin of the Green anole (*Anolis carolinensis*). We report 1,719 differentially expressed genes 23 of which are involved in chromatophore development, biosynthesis of melanin and pteridine pigments, or carotenoid transport and sequestration. In the dewlap, transcripts of the xanthophore associated gene (*csflr*) and three genes possibly involved in carotenoid transport and sequestration (*cd36*, *gsta2*, *plin2*) are significantly up-regulated. Because pteridine pigments are known to contribute to dewlap coloration in *A. carolinensis*, we also use quantitative PCR to investigate whether genes in the pteridine pigment biosynthetic pathway are differentially expressed as well as whether pteridine gene expression varies ontogenetically. With both qPCR and

transcriptome sequencing approaches we find that *spr*, a key gene in the production of the yellow sepiapterin pigment, is more highly expressed in the dorsal skin than in the dewlap or ventral skin, and that the pteridine genes are expressed at constant levels independent of ontology. These results identify a set of candidate genes regulating the distribution of carotenoid and pteridine pigments in anole integument. The genes involved are of broad interest as their function is evolutionarily conserved in both insects and vertebrates and thus they are likely to be generally involved in colorful pigmentation of vertebrates .

Introduction

A key challenge in evolutionary biology is understanding the genetic basis of traits that play important roles in local adaptation and speciation. Coloration is central to the biology of most macroorganisms. Cryptic coloration provides camouflage from predators and is typically under strong ecological selection (reviewed in: Bond 2007). Conspicuous coloration is widespread in the animal kingdom and can be seen on the wings of butterflies, the plumage of many species of birds, the patterns of some fishes, and even the rumps and faces of some primates (Bradley & Mundy 2008; Protas & Patel 2008; Trapnell *et al.* 2010; Kronforst *et al.* 2012; Greenwood *et al.* 2012; Kim *et al.* 2013). Conspicuous coloration may function as a warning to predators, or as a social signal used for territorial or mating displays (reviewed in: Baynash *et al.* 1994; Dupin & Le Douarin 2003; Mills & Patterson 2009; Zambon *et al.* 2012) and coloration plays a central role in sexual selection. Surprisingly, although the genetics of cryptic coloration have been well studied in vertebrates, the genetic basis for conspicuous coloration is less

well understood (reviewed in: Hubbard *et al.* 2010).

Anolis lizards are an exceptional system in which to study the genetics of conspicuous coloration because the males of most species have brightly colored extendable throat fans called dewlaps that are used in social encounters (**Figure 2-1**) (Jenssen 1977). Dewlap coloration is spectacularly variable among anole species and includes reds, oranges, blues, greens, blacks, and whites which are found on dewlaps in a diversity of patterns and arrangements (Nicholson *et al.* 2007). Anoles also display an array of cryptic dorsal colors including browns, greens, blues, yellows, and greys (Paemlaere *et al.* 2011). Additionally, anoles are an excellent system in which to study coloration because they represent one of the largest vertebrate radiations (> 400 species), have a well resolved phylogeny, and a complete genome is available for one species (*Anolis carolinensis*) (Alföldi *et al.* 2011). Finally, similar coloration appears to have evolved independently in multiple species providing the opportunity to examine the genetic basis of convergent evolution in coloration.

In this study, we investigate the genetic components of coloration by examining patterns of gene expression in differently colored skin tissue in the green anole (*A. carolinensis*). *Anolis carolinensis* is the only anole native to the continental U.S. where it has a wide distribution across the southeastern states (Tollis *et al.* 2012; Campbell-Staton *et al.* 2012). In *A. carolinensis* the dewlap is pink, the dorsum green, and the ventral skin is white (**Figure 2-1**). The pink dewlap contains both red and yellow pigments, melanocytes, and iridophores that produce blue, and potentially ultraviolet, structural hues. How the pink color is produced is not precisely known. The green color

of the dorsum is produced by a combination of yellow pigments (both pteridine and carotenoids are present) and blue structural colors (Alexander & Fahrenbach 1969). The blue structural colors are not pigments, but are instead produced by specialized cells (iridophores) that contain guanine platelets that refract and reflect blue light (Alexander & Fahrenbach 1969) but see Rohrllich & Rubin (1975). Making the system additionally complex, while the dorsal skin is typically green, it also contains melanophores that have extensions towards the surface of the dermis which allow melanin to be transported upward, giving the animal a brown appearance when stimulated by melanocyte-stimulating hormone (Horowitz 1957; Goldman 1969).

Red and yellow colors in *A. carolinensis* result from two types of pigments: pteridines and carotenoids (Ortiz & Maldonado 1966; Macedonia *et al.* 2000). Pteridine pigments are synthesized from guanosine-5'-triphosphate (GTP), while carotenoid pigments are sequestered from the diet (primarily insects). About ten genes, first identified in zebrafish and *Drosophila*, are involved in the production of red and yellow pteridine pigments (Forrest & Mitchell 1954; Ziegler *et al.* 2000; Ziegler 2003; Braasch *et al.* 2007). However, the mechanisms regulating the sequestration and deposition of carotenoid pigments are almost entirely unknown, and only a few genes associated with carotenoid mutants have been identified (Walsh *et al.* 2012).

Here, we investigate patterns of gene expression in the white ventral, green dorsal, and pink dewlap skin of *A. carolinensis* (Figure 2-1), using both qPCR of candidate genes and analysis of mRNA transcriptome profiles, to identify genes associated with the different coloration of the skin. We first analyze the complete

transcriptomes of these three tissues to identify a subset of genes potentially involved in pigmentation. Because pteridine pigments are known to be important components of pigmentation in *A. carolinensis* (Ortiz *et al.* 1963; Macedonia *et al.* 2000; Olsson *et al.* 2013) we then use qPCR to examine expression differences in genes involved in the pteridine pathway among tissues. We also investigate whether pteridine gene expression varies ontogenetically, in juvenile through adult lizards, to determine if these genes are active only during the period when sexual maturity and adult coloration develops in males.

Results

We sequenced 9.3, 8.3, and 3.2 million reads from the dewlap, dorsal and ventral skin transcriptomes of ten adult *A. carolinensis* males, for a total of 20.9 million reads (Methods: Pooled RNA sequencing). We calculated Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values using the Ensembl gene predictions (vsn. 67) to identify differentially expressed genes. Although there are more than twice as many reads in the dewlap and dorsal skin relative to ventral skin, counts of genes with non-zero FPKM values are similar across tissues (Dewlap = 14,917; Dorsal = 14,609; Ventral = 13,165). Furthermore 10,562 of these gene ids are shared between all three tissues, which demonstrates that each tissue is expressing a similar set of genes. In pairwise comparisons between tissues, we identified 1,719 significant differences in FPKM between unique genes at a false discovery rate (FDR) of 0.05 (**Table 2-1, Supp Figure 2-1; Supp. Figure 2-2**). These genes represent 8.1% of annotated genes.

Gene Ontology terms.

As a first step in identifying genes and pathways contributing to the color of the skin, we ran a gene ontology (GO) analysis to investigate the functions of genes whose transcripts were differently expressed in the three tissues. This analysis identified 60 GO terms enriched in the dewlap and 53 GO terms enriched in the dorsal skin (p-value < 0.05; FDR = 0.05; **Supp. Table 2-1**), but no GO terms were significantly enriched in the ventral skin. In the dorsal skin GO terms associated with melanin (GO:0006582), melanocyte differentiation (GO:0030318), pigment granule (GO:0048770), and pigmentation (GO:0043473) were significantly overrepresented. Additionally, GO terms associated with neural crest cell migration and development (GO:0001755, GO:0014032) and epithelial growth and development (GO:0045682, GO:0050680) were also overrepresented. Chromatophores (pigment containing cells) are derived from neural crest cells (Dorsky *et al.* 1998; Ziegler *et al.* 2000; Dupin & Le Douarin 2003) and contain pteridine and carotenoid pigments (Alexander & Fahrenbach 1969; Ortiz & Maldonado 1997; Macedonia *et al.* 2000). The dorsal skin also contains a substantial number of melanocytes, as many as 50 per scale (Alexander & Fahrenbach 1969), which enable facultative color change from green to brown (Goldman 1969).

In the dewlap, the GO analysis did not identify distinct pigmentation pathways. However, dewlap genes associated with purine ribonucleoside metabolic processes (GO:0046128) are significantly overrepresented. Pteridine pigments are constructed from purines, so the pathway has potential biological significance. Two lysosome associated GO terms (GO:0007040, GO:0005764) are also overrepresented. Lysosomes are closely

related to pteridine and melanin containing organelles (Lloyd *et al.* 1998; Navarro *et al.* 2008) so genes in this pathway may be contributing to coloration. We found no carotenoid associated GO terms to be overrepresented, but ‘cellular lipid catabolic process’ (GO:0044242) could be involved in the metabolism and transport of carotenoids which are transported as low-density lipoproteins (Walsh *et al.* 2012).

In addition to the GO analysis, we compiled a list of 75 candidate genes from the literature (**Supp. Table 2-2**). This list includes genes associated with melanin, pteridine, and carotenoid pigmentation as well as genes associated with iridophores and keratin proteins (potential producers of structural colors: Prum *et al.* 2009). We searched this list for significant differences in expression between dermal tissues. One complicating factor in this analysis is that the standard deviation around FPKM values is about two-fold larger in the ventral skin because of the relatively small number of reads sequenced for this sample (standard deviation in FPKM: Dorsal = 2,676.56; Dorsal = 2,909.74; Ventral = 3,973.24). This reduces the power to detect differences in comparisons with ventral skin. Still, we found that 35 genes in this list are differentially expressed (p-value < 0.05; FDR = 0.05; Table 2-2).

Melanin associated genes.

The melanin pathway is well-characterized, both because of its importance as a target of ecological selection (Hoekstra 2006), and because melanin containing cells (melanocytes) are associated with melanoma (Franck & Schatton 2008; Bandarchi *et al.*

2010). The melanin pathway can produce two types of pigments, eumelanin and pheomelanin, but lizards only produce eumelanin (Bagnara & Hadley 1973). The eumelanin pathway consists of four genes: tyrosinase (*tyr*), tyrosinase-related protein 1 (*tyrp1*), dopachrome tautomerase (*dct*), and silver (*silv* or *pmel*) which are members of the tyrosinase gene family (reviewed in: Hearing & Tsukamoto 1991; Braasch *et al.* 2007). As expected, transcripts of these genes, except for *dct*, are significantly more abundant in the dorsal skin.

Similarly, transcripts of the melanosomal transporters *oca2* and *slc24a5* are significantly more abundant in the dorsal skin relative to both the dewlap and the ventral skin. Additionally, transcripts of the melanosome associated, seven transmembrane G-protein coupled receptor (*endrb*), as well as a similar melanosomal receptor (*endrb2*) are both more abundant in the dorsal skin. Transcripts of the ‘absent in melanoma 1-like protein’ (*aim1*) were significantly less abundant in the dorsal skin relative to both the dewlap and ventral skin. Phosphodiesterase 4 (*ped4*) regulates melanosome dispersal during environmentally regulated color change in the skin of zebrafish (*Danio rerio*) (Richardson *et al.* 2008). Intriguingly, transcripts of *ped4* were more abundant in the dewlap, but not in the dorsal skin. One explanation for this observation could be that while *ped4b* is the closest homologue to *ped4* in the green anole, it may not have the same function as *ped4* in zebrafish. Taken together the GO analysis and the candidate gene analysis suggest that the melanin pathway is up-regulated in the dorsal skin. This is congruent with *A. carolinensis* dorsal skin tissue containing large numbers of melanophores and melanin pigment because it contains the necessary cellular apparatus

for facultative color change from green to brown.

Pteridine associated genes.

The pteridine pathway is active in anoles, as is demonstrated by a diversity of pteridine pigments that can be extracted from the dewlap and body skin (Ortiz & Maldonado 1966; Macedonia *et al.* 2000). However, we only identified one pteridine gene, aldehyde oxidase 1 (*aox1*) whose transcript profile varied among tissues. In zebrafish, aldehyde oxidase 3 (*aox3*) is involved in xanthoblasts, a cellular precursor to xanthophore cells (Parichy *et al.* 2000; Curran *et al.* 2010). Xanthopterin is a yellow pigment so it is possible that *aox1* contributes to the production of yellow pigments. But it is surprising that transcripts of other genes in the pteridine biosynthetic pathway were not more abundant in the dewlap relative to other tissues given the demonstrated presence of pteridine pigments in the dewlap (Ortiz *et al.* 1963; Ortiz & Maldonado 1966; Macedonia *et al.* 2000).

The dewlap of *A. carolinensis* contains red-orange drosopterin pigments (Ortiz & Maldonado 1966). In *Drosophila*, the *clot* gene, which is possibly homologous to thioreductase-like 5 (*txn15*) in vertebrates, appears to regulate the production of drosopterins (Braasch *et al.* 2007). However, *txn15* is not annotated in the anole genome and blast searches of *clot* and *txn15* homologs from *Drosophila* and zebrafish against the *A. carolinensis* genome assembly do not produce hits with low e-values. There are 22 thioreductases annotated in the anole genome, however none of these gene transcripts are significantly differently abundant in the dewlap. Thus we were unable to find a gene

homologous to the *Drosophila clot* gene, and the gene or genes responsible for drosoplerin production in the dewlap remain elusive.

Pteridine pigments are typically contained within xanthophores and the colony stimulating factor 1 receptor (*csf1r*) gene has been associated with the development of pteridine containing xanthophores (Rousset *et al.* 1995; Parichy *et al.* 2000; Salzburger *et al.* 2007). We observed that transcripts of *csf1r* are more abundant in the dewlap relative to the ventral skin. The dorsal versus ventral comparison, however, is not significant and is likely due to the relatively low number of reads and high variance in transcript abundance for the ventral skin.

Carotenoid associated genes.

The sequestration and deposition of carotenoid pigments are not well understood in vertebrates (Hearing & Tsukamoto 1991; McGraw & Toomey 2010). However, recent work has identified 11 genes associated with carotenoid pigmentation in insects, crustaceans, birds, and mammals (reviewed in: Hearing & Tsukamoto 1991; Walsh *et al.* 2012). Two of these genes are differentially expressed in *A. carolinensis* dewlap. Transcripts of *cd36* (also known as *SCARB3*) are more abundant in dewlap relative to the dorsal skin, and transcripts of perilipin 2 (*plin2*) are more abundant in the dewlap relative to the ventral skin. *Cd36* is a putative transmembrane lipoprotein receptor that regulates carotenoid uptake in silkworms (*Bombyx mori*) (Sakudoh *et al.* 2010; 2013). Mutations in *cd36* are also important indicator of malaria susceptibility in human populations (Aitman *et al.* 2000), which suggests that *cd36* could be important in parasite resistance in anoles.

In the Guadeloupean anole (*Anolis marmoratus*), *cd36* appears to be under divergent ecological selection (Chapter 3: this dissertation). *Plin2* is involved in packaging or breakdown of lipid droplets within cells of adipose tissue and may be associated with carotenoid transport and sequestration (Walsh *et al.* 2012). Concordantly, neither of these genes is differentially expressed in the dorsal skin which should contain mostly pteridine pigments (Ortiz & Maldonado 1966). Together this suggests a role for *cd36* and *plin2* in the deposition of carotenoid pigments in the dewlap of *A. carolinensis*.

Genes associated with blue structural colors.

Blue coloration in anoles is thought to be produced by iridophore cells, which contain reflective guanine platelets (Alexander & Fahrenbach 1969). However, Rohrlich and Rubin (1975) suggest that the guanine platelets do not reflect blue light. This leaves a potential role for keratins in producing blue color as keratin nanostructures have been shown to produce blue structural colors in birds (Prum *et al.* 2009). Transcripts of Purine Nucleoside Phosphorylase 4a (*pnp4a*), an iridophore marker gene (Curran *et al.* 2010), are highly abundant in the dewlap. This is concordant with a histological studies that find that the dewlap contains a dense layer of iridophores (Alexander & Fahrenbach 1969; Hearing & Tsukamoto 1991). Anole integument also contains a diversity of keratin genes (Alibardi *et al.* 2012) seven of which have been annotated in Ensembl (**Supp. Table 2-2**).

Of these genes, two have transcripts that are more abundant in dewlap and dorsal skin. In the dewlap and ventral skin, transcripts of keratin 80 (*krt80*) are more abundant relative to the dorsal skin, while transcripts of keratin 18 (*krt18*) are more abundant in the dorsal skin relative to the dewlap. While these differences may be associated with

structural differences in the scales of the different tissues, the potential role of keratin ultrastructure in producing blue coloration in anoles deserves further study.

qPCR of pteridine genes.

Pteridines are known to be abundant pigments in anole skin (Ortiz & Maldonado 1966; Macedonia *et al.* 2000), but our transcriptomic analyses only found one pteridine related gene (*aox1*) to be differentially regulated. One potential explanation of this observation is that genes producing pteridine pigments may be active only during ontogeny when adult male coloration develops. To investigate the pteridine pathway in more detail we used quantitative PCR (qPCR) to measure pteridine expression in 20 male *A. carolinensis* representing an ontogenetic series from sub-adults (SVL < 40mm) to large adult males (SVL > 60mm). We interrogated three pteridine genes: 6-pyruvoyltetrahydropterin synthase (*pts*), sepiapterin reductase (*spr*), and ‘protein associated with Myc’ (*pam*).

Pts regulates the synthesis of 6-pyruvoyl-tetrahydropterin, the precursor compound for both red-orange drosopterin and yellow sepiapterin pigments. Differential expression of *pts* may contribute to gross changes in pigmented versus unpigmented skin (i.e., ventral vs dorsal and dewlap). *Spr* regulates the production of yellow sepiapterin (Ziegler *et al.* 2000; Ben *et al.* 2003) and *pam* is also thought to play a role in the regulation of yellow sepiapterin pigments (Le Guyader *et al.* 2005). Therefore we expected *spr* and *pam* to be more highly expressed in the dorsum than other tissues. As noted previously, we were unable to identify a homolog of the *Drosophila* *clot* gene in the *A. carolinensis* genome and transcriptome so were unable to interrogate the pathway

leading to the production of drosopterin, likely an important component of red pigmentation in the dewlap.

To investigate whether transcripts of genes in the pteridine pathway were differentially abundant, we performed an analysis of covariance (ANCOVA) on the $\Delta\Delta Ct$ values from the qPCR analysis (Methods: Quantitative PCR; **Table 2-3**).

The ANCOVA for gene versus tissue was significant at a p-value < 0.05 [$F(2) = 4.92$; $p = 0.0089$] which indicates that the pteridine pathway is differentially regulated in one or more tissues. Snout-vent length, however, did not significantly co-vary with gene expression [$F(2) = 0.65$; $p = 0.52$]. This indicates that ontogeny, as indexed by SVL, does not affect expression of the pteridine pathway. To determine if any gene transcripts were significantly more abundant in any tissue, we performed a multivariate analysis of covariance (MANCOVA) of each gene versus tissue (Table 2-4). The MANCOVA found that *spr* was significantly different between tissues [$F(1) = 13.069$, $p = 0.00086$] and *pam* was marginally significant [$F(1) = 2.63$, $p = 0.113$]. *Spr* transcripts are more abundant in dorsal skin relative to the dewlap (Table 2-4), consistent with the expectation that yellow sepiapterin pigment is more abundant in the dorsum where the green color is produced by a combination of yellow pigment and blue structural coloration (Alexander & Fahrenbach 1969).

Discussion

This study is the first to comprehensively investigate the expression of genes involved in skin pigmentation in anoles. In our comparison of three transcriptomes from dewlap, dorsal, and ventral skin we identified 1,719 significant pairwise comparisons

between genes in the *A. carolinensis* genome. These genes include genes in melanin, pteridine, and carotenoid pigmentation pathways. Both the gene ontology analysis (GO) and our search for significantly expressed candidate genes identified genes involved in the melanin pathway as being abundant in the dorsal skin. We also identified genes that contribute to the conspicuously colored dewlap and to the green dorsal skin. The genetic basis of conspicuous coloration is poorly understood in vertebrates. However, one gene, *csflr*, is associated with conspicuous coloration in cichlid fish and has been shown to play an important role in sexual selection (Parichy *et al.* 2000; Salzburger *et al.* 2007). We also identify *csflr* as an important contributor to anole coloration. Additionally, two genes, *cd36* and *plin2*, that are potentially associated with carotenoid pigmentation, appear to play important roles in the pigmentation of the dewlap in *A. carolinensis*.

Colony stimulating factor 1 receptor (*csflr*) regulates the development of yellow pigment containing xanthophores in zebrafish (Parichy *et al.* 2000). In some species of cichlid fish *csflr* contributes to the coloration of sexually selected egg-dummy spots on the anal fins of males, and there is molecular evidence that portions of *csflr* are positively selected (Salzburger *et al.* 2007). Histological evidence suggests that the dewlap contains a special type of chromatophore called an erythrophore which contains only red pigments (Alexander & Fahrenbach 1969). So, the finding that *csflr* transcripts are most abundant in the dewlap and not in the dorsum suggests that *csflr* may be associated with erythrophore development in *A. carolinensis*.

The genes involved in carotenoid pigmentation are not well understood because the pathway of carotenoid metabolism, transport and sequestration involves a multitude of tissues: carotenoids are first absorbed in the gut, then processed and packaged in the liver, transported in the blood, and finally deposited in chromatophores. Cluster of differentiation 36 (*cd36*) functions as a transmembrane lipoprotein receptor that may be involved in the uptake of carotenoids (Sakudoh *et al.* 2010; Sakudoh *et al.* 2013).

Perilipin 2 (*plin2*) coats lipid storage droplets and may play a role in carotenoid storage (Londos *et al.* 1996; Walsh *et al.* 2012). Our finding that transcripts of these two genes are significantly more abundant in the pink skin of the dewlap, which is known to contain carotenoid pigments (Ortiz *et al.* 1963; Macedonia *et al.* 2000; Steffen & McGraw 2007) suggests that they may be contributing to the deposition and storage of carotenoid pigments.

Analysis of the ontogenetic series revealed that pteridine pathway genes are continuously expressed throughout ontogeny. This is not surprising since pteridines may play multiple functional roles and they continuously undergo oxidative degradation and so require replacement. The finding that *spr* and *aox1* are up-regulated in the dorsum relative to the dewlap is consistent with the production of yellow pteridine pigments in the dorsal skin where they contribute to the green coloration of the skin. However, pteridine pathway genes involved in producing the red-orange drosopterin pigment, which is present in the dewlap of *A. carolinensis* (Ortiz & Maldonado 1966; Macedonia *et al.* 2000), remain elusive. In *Drosophila*, the *clot* gene is responsible for production of

drosoplerin pigment in the orange eyes of wild-type *D. melanogaster*. While we failed to find a *clot* homolog in the annotated *A. carolinensis* genome, it should be noted that, in our transcriptomic analysis of *A. carolinensis* tissues, we identified a small, 180bp transcript that was a possible homolog to the *Drosophila clot* gene. This transcript was present only in the dewlap, as would be predicted for a gene involved in drosoplerin production. However, qPCR using primers designed from this fragment did not reveal differential expression among tissues. Thus we do not have confidence that this fragment represents a functional copy of a *Drosophila clot* homolog.

Dewlap pigmentation is of great interest from an evolutionary perspective (Nicholson *et al.* 2007). Anoles use their dewlaps in social signaling displays (Jensen 1977), and one hypothesis is that the color of the dewlap is an honest indicator of quality (Losos 2009). If this is the case, carotenoid pigments should contribute substantially to the color of the dewlap because carotenoids are obtained from dietary sources and play important roles in health (Hill *et al.* 2002; Clotfelter *et al.* 2007; Baeta *et al.* 2008). Ortiz *et al.* (1963) found that carotenoid pigments were abundant in yellow-pigmented dewlaps and likely also contribute to orange colored dewlaps. Thus *cd36* and *plin2* are potentially important genes involved in carotenoid pigmentation. Because the dewlap is an important component of male-male and male-female interactions, the *cd36* and *plin2* genes may be evolving under sexual selection. Comparative analysis of these genes in polymorphic taxa and across species will be informative to determine if there are molecular signatures of natural selection.

From a broader perspective, changes in pigmentation may drive speciation in anoles. In anoles, differences in pigmentation appear to occur prior to the evolution of larger morphological changes (e.g., ecomorph evolution) (Losos 2009) and divergence in coloration due to ecological or sexual selection may lead directly to reproductive isolation. Thus, the genes and pathways involved in pigmentation may be critical ‘speciation genes’ or ‘speciation pathways’ whose divergent evolution pleiotropically contributes to reproductive isolation. Future work investigating molecular evolution and gene expression in *csflr*, *cd36*, and *plin2* will help to illuminate the relative roles of these genes both in the context of sexual selection as well as in the context of the adaptive radiation of *Anolis* lizards.

Methods

Phenotypic measurements.

We collected photos, color spectra, and size measurements for all lizards used in this study. We photographed the dorsal, ventral and dewlap of each lizard against a white one-centimeter grid alongside a Munsell ColorChecker Mini color-standard card. We measured color-spectra with an Ocean Optics USB 4000 field-portable spectrometer. We recorded reflectance values as percent reflectance relative to a barium sulphate white standard using an Ocean Optics UV-VIS reflectance probe attached to a PX-2 pulsed xenon UV-VIS light source (Ocean Optics, Inc., Dunedin, FL, USA). Snout-Vent length (SVL) for each lizard was measured to the nearest 0.1 centimeter with digital calipers.

Pooled RNA-sequencing.

To sequence the transcriptomes of dewlap, dorsal, and ventral skin we dissected fresh skin from 10-11 adult male *A. carolinensis* into 1 mL of TRIzol®, homogenized the 5-10 mg of tissue through repeated pumping with a syringe and needle, and extracted total RNA following the TRIzol® instructions. We purified mRNA with oligo-dT beads (Dynabeads®). mRNA purification from samples of each tissue from single lizards did not produce enough material to construct libraries (e.g., < 1µg total RNA). Three RNA libraries were constructed following modified standard Illumina methods (**Supp. Method 2-1**). Libraries were individually tagged with a CCT, AAT, or GGT inline tag (Craig *et al.* 2008; Peterson *et al.* 2012). Libraries were pooled and 36 bp reads were sequenced on two lanes of an Illumina GAI.

We mapped reads to the *A. carolinensis* genome (anoCar2, Ensembl version 2.0.67) with Tophat (version, 2.0.4) and assessed differential transcript abundance with Cufflinks (version, 2.0.2) (Trapnell *et al.* 2010; Kim *et al.* 2013). To remove technical artifacts we applied upper quartile normalization (Irizarry *et al.* 2003) as well as correcting for fragment bias due a non-uniform distribution of read sequences (e.g., primer contamination) and reads mapping to multiple genomics locations.

Overrepresented gene ontologies based on the presence or absence of genes were identified with GO-Elite (Zambon *et al.* 2012) at an FDR of 0.05. Subsequent statistical analyses were preformed with a combination of the python data analysis library (<http://pandas.pydata.org/>), R statistical software (R Core Team 2012), and custom python scripts.

Single Sample RNA sequencing.

We also sequenced the complete transcriptome of one dewlap from a single adult male. RNA was extracted from fresh tissue using the trizol method described above, but libraries were constructed using an Illumina kit and un-tagged paired-end primers were used. The library was sequenced on one 101bp lane of an Illumina HiSeq.

Quantitative PCR.

We used TRIzol® and syringe and needle homogenization to extract total RNA individually from dewlap, dorsal, and ventral skin of twenty-two male *A. carolinensis*. These lizards represent an ontogenetic series from juvenile to adult and ranged in size (SVL) from 33.6 - 61.8 mm (**Figure 2-3**). Total RNA was extracted with TRIzol® and converted to cDNA using oligo-dT primers and superscript reverse transcriptase II. Taqman® assays were performed on an Eppendorf Realplex² Mastercycler for each sample for the following genes in the pteridine synthesis pathway:

6-pyruvoyltetrahydropterin synthase (*pts*), sepiapterin reductase (*spr*), protein associated with Myc (*pam*). β -actin (*actb*) was used as an internal control to account for differences in the quantity of cDNA in each reaction. Probe sequences were designed from ensembl gene predictions and are reported in **Supp. Table 2-3**.

Initial analyses revealed that many transcripts could not be mapped to the AnoCar2 assembly so, to improve our ability to map transcripts and identify genes, we produced a complete transcriptome of the dewlap. We sequenced one lane of 101 bp paired-end reads on an Illumina HiSeq. These reads were assembled *de novo* with ABySS (version 1.3.3) at K=58 (Simpson *et al.* 2009). We used tblastx to search for clot

homologs within the denovo assembly. Two overlapping hits (e-values $> 1 \times 10^{36}$) were aligned with the MUSCLE sequence aligner (<http://www.ebi.ac.uk/Tools/msa/muscle/>). This 180 base-pair fragment reciprocally blasts (e.g., blast searches using the clot gene from *Drosophila* versus the transcriptome and vice versa always return a *clot* gene) to thioredoxin domains (both e-values are $\sim 9 \times 10^{26}$). Primer3 (v1.14) was used to design intron-spanning primers. cDNA was prepared as previously described. However, a KAPA SYBR® FAST qPCR kit was used to assay gene expression instead of Taqman® assays.

For all qPCR reactions, we assayed each sample for each gene in duplicate. When the difference in PCR cycle threshold (Ct) between replicates was greater than one, indicating a potential pipetting or PCR error, both measures were discarded. We compared expression between tissues and genes using an efficiency-calibrated model of normalized gene expression ($\Delta\Delta\text{Ct}$). We calculated $\Delta\Delta\text{Ct}$ using β -actin as an internal reference gene and the unpigmented ventral sample as the calibrator (Applied Biosystems User Bulletin No. 2 (P/N 4303859) and reviewed in: (Livak & Schmittgen 2001).

Figures and Tables

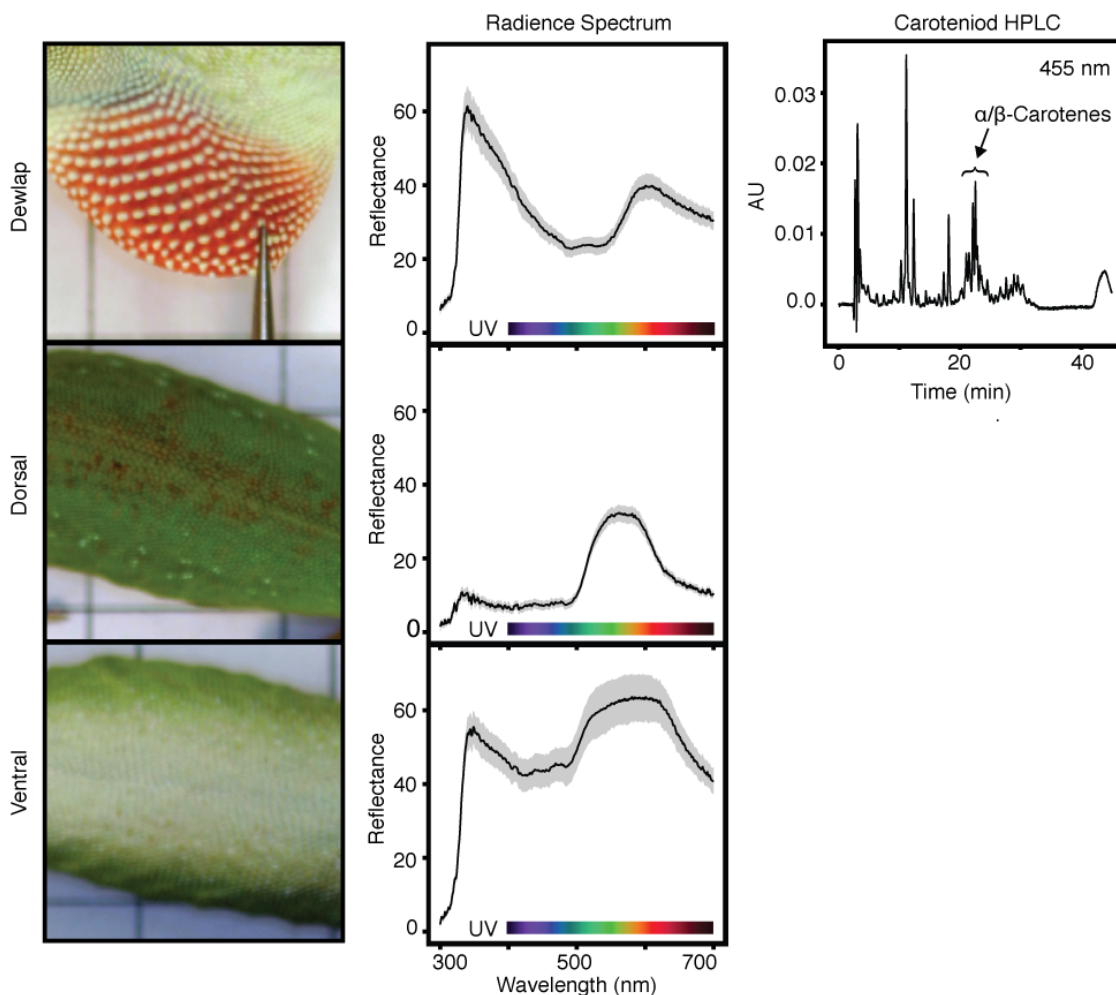


Figure 2-1. The first column depicts dewlap, dorsal, and ventral skin from the green anole (*Anolis carolinensis*). The second column shows the radiance spectrum from 300-700 nanometers. Spectral measurements from 22 lizards smoothed with a 100 nm hamming window. The grey ribbon represents a one standard deviation around the mean. RGB colors in the visible spectrum are included for reference. The last column shows HPLC measurements for carotenoid pigments including α and β carotenes.

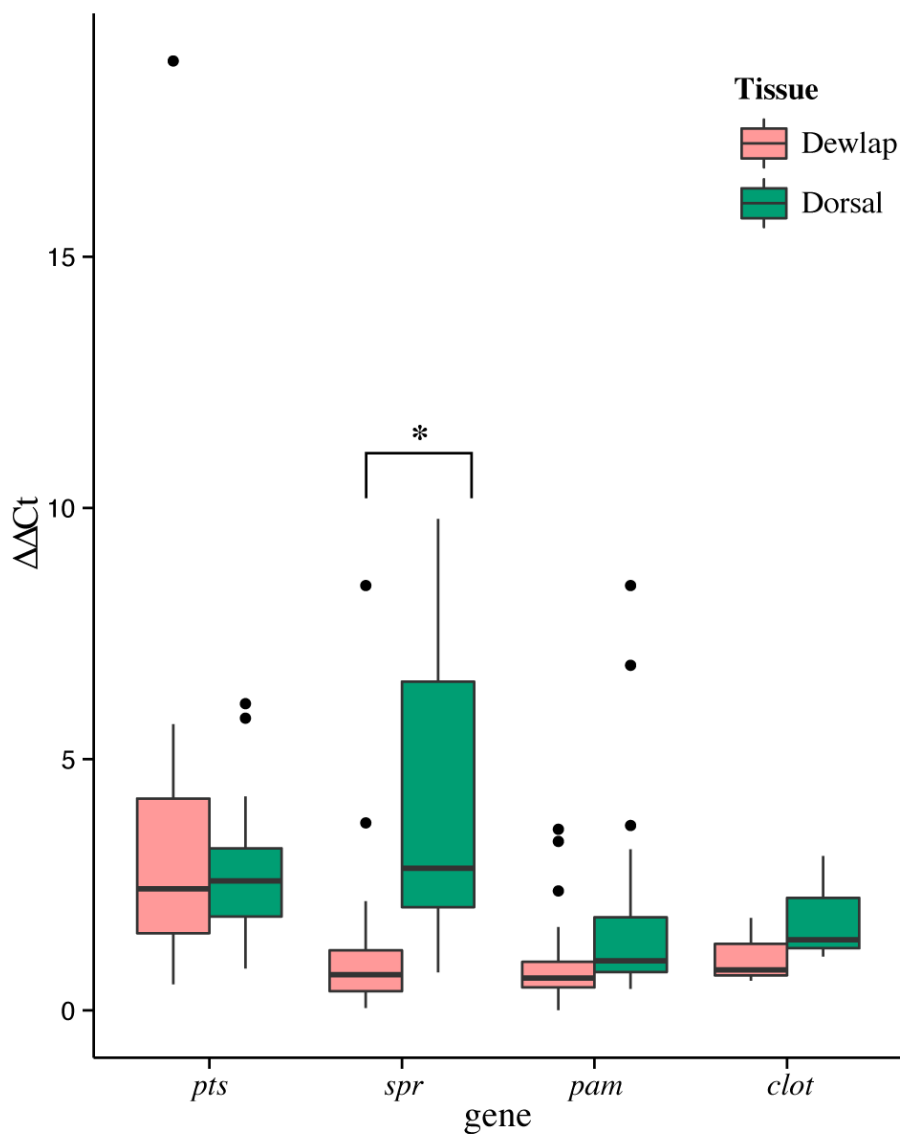


Figure 2-2. Gene expression ($\Delta\Delta Ct$) for *pts*, *spr*, and *pam* is from 22 lizards ranging in size from 33.6 mm to 61.77 mm (mean = 49.18 mm \pm 9.32 stdev). Gene expression for *clot* is from 5 lizards ranging in size from 57.1mm to 61.6mm (mean = 59.7 \pm 1.8 stdev). $\Delta\Delta Ct$ is normalized expression of each gene versus β -actin with the ventral skin used as a calibrator. * indicates significant differences ($p < 0.05$) (see Table 2-4).

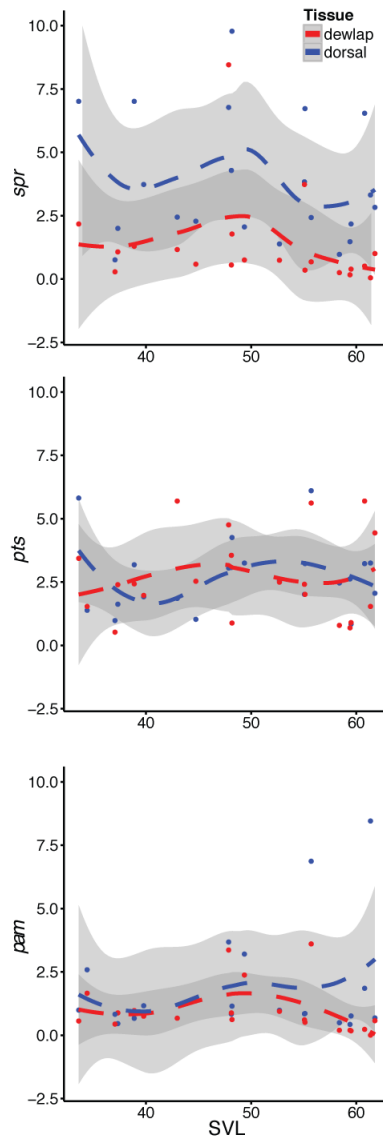


Figure 2-3 Scatter plot of *spr*, *pts*, and *pam* gene expression from 22 lizards ranging in size from 33.6 mm to 61.77 mm. $\Delta\Delta Ct$ is gene expression is normalized expression of each gene vs β -actin (= housekeeping gene) and the ventral skin (=calibrator). The dashed fit line is LOESS smoothed. The grey bands represent 95% confidence intervals.

Tables

	Dewlap	Dorsal	Ventral
Dewlap	-	101	47
Dorsal	580	-	486
Ventral	321	184	-

Table 2-1 Counts of genes that are significantly down or up regulated in dewlap, dorsal, and ventral skin in the RNA-seq analyses. The upper triangle contains the counts of significantly up-regulated genes while the lower triangle contains the counts of significantly down-regulated genes. Significance was determined with Cuffinks2 at an FDR of 0.05. Only those genes included in the Ensembl 67 *A. carolinensis* Gene Predictions were included in the calculations.

Pathway/ Cell	Gene name	Tissue 1	Tissue 2	FPK M 1	FPK M 2	Fold change	P value	Q value
Carotenoid	<i>cd36</i>	Dewlap	Dorsal	68.72	22.61	-1.60	0.001	0.016
Carotenoid	<i>gsta2</i>	Dewlap	Ventral	52.69	180.46	1.78	< 0.001	0.006
Carotenoid	<i>plin2</i>	Dewlap	Ventral	27.29	9.36	-1.54	0.002	0.047
Iridophore	<i>pnp</i>	Dewlap	Dorsal	62.78	0.68	-6.52	< 0.001	< 0.001
Iridophore	<i>pnp</i>	Dorsal	Ventral	0.68	44.08	6.01	< 0.001	< 0.001
Keratin	<i>krt18</i>	Dewlap	Dorsal	9.98	35.00	1.81	< 0.001	0.008
Keratin	<i>krt80</i>	Dewlap	Dorsal	53.97	9.46	-2.51	< 0.001	< 0.001
Keratin	<i>krt80</i>	Dorsal	Ventral	9.46	52.40	2.47	< 0.001	< 0.001
Melanin	<i>pde4b</i>	Dewlap	Dorsal	15.98	5.16	-1.63	0.001	0.034
Melanin	<i>tyrp1</i>	Dewlap	Dorsal	0.73	14.52	4.31	< 0.001	< 0.001
Melanin	<i>tyrp1</i>	Dorsal	Ventral	14.52	0.35	-5.39	< 0.001	< 0.001
Melanin	<i>pmel</i>	Dorsal	Ventral	49.01	4.67	-3.39	< 0.001	< 0.001
Melanin	<i>pmel</i>	Dewlap	Dorsal	10.25	49.01	2.26	< 0.001	< 0.001
Melanin	<i>tyr</i>	Dewlap	Dorsal	0.36	4.53	3.65	< 0.001	0.008
Melanin	<i>tyr</i>	Dorsal	Ventral	4.53	0.60	-2.93	< 0.001	0.008
Melanin	<i>tyrp1</i>	Dewlap	Dorsal	0.73	14.52	4.31	< 0.001	< 0.001
Melanin	<i>tyrp1</i>	Dorsal	Ventral	14.52	0.35	-5.39	< 0.001	< 0.001
Melanin	<i>ednrb2</i>	Dorsal	Ventral	13.79	0.88	-3.98	< 0.001	< 0.001
Melanin	<i>ednrb2</i>	Dewlap	Dorsal	2.76	13.79	2.32	< 0.001	0.004
Melanin	<i>ednra</i>	Dorsal	Ventral	9.16	1.12	-3.03	< 0.001	0.005
Melanin	<i>ednra</i>	Dewlap	Dorsal	1.30	9.16	2.82	< 0.001	0.006
Melanin	<i>aim11</i>	Dewlap	Dorsal	38.98	5.62	-2.79	< 0.001	< 0.001
Melanin	<i>aim11</i>	Dorsal	Ventral	5.62	41.18	2.87	< 0.001	< 0.001
Melanin	<i>oca2</i>	Dorsal	Ventral	2.20	0.00	NA	< 0.001	0.015
Melanin	<i>slc24a5</i>	Dewlap	Dorsal	0.47	4.89	3.37	0.001	0.018
Melanin	<i>slc24a5</i>	Dorsal	Ventral	4.89	0.79	-2.63	0.001	0.025
Melanin	<i>cdh11</i>	Dorsal	Ventral	8.25	2.38	-1.80	0.001	0.029
Melanin	<i>plcb2</i>	Dewlap	Dorsal	4.53	1.16	-1.96	0.001	0.016
Melanin	<i>plcb4</i>	Dewlap	Ventral	3.09	0.77	-2.01	0.001	0.030
Pteridine	<i>prdx1</i>	Dorsal	Ventral	45.81	135.27	1.56	0.001	0.027
Pteridine	<i>prdx1</i>	Dewlap	Dorsal	130.76	45.81	-1.51	0.001	0.034
Pteridine	<i>txnrd1</i>	Dorsal	Ventral	17.93	55.45	1.63	< 0.001	0.015
Pteridine	<i>aox1</i>	Dewlap	Dorsal	2.57	9.20	1.84	< 0.001	0.010
Xanthopore	<i>csflr</i>	Dewlap	Ventral	11.99	3.59	-1.74	0.001	0.022

Table 2-2 Significantly differentially expressed pigmentation genes from the transcriptomic analyses (p-value = 0.05; FDR=0.05). Ensembl id and chromosomal position are reported in (Supp. Table 2-2).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Gene	2	62.49	31.24	5.65	0.0046
Tissue	1	23.47	23.47	4.24	0.0416
SVL	1	0.08	0.08	0.01	0.9044
Gene:Tissue	2	54.42	27.21	4.92	0.0089
Gene:SVL	2	7.19	3.6	0.65	0.5237
Tissue:SVL	1	1.26	1.26	0.23	0.6336
Gene:Tissue:SVL	2	3.39	1.69	0.31	0.7367
Residuals	116	641.5	5.53		

Table 2-3 Analysis of covariance (ANCOVA) comparison between tissues and genes.

Gene		Df	Sum Sq	Mean Sq	F value	Pr(>F)
PTS	Tissue	1	5.76	5.764	0.6446	0.427
	Residuals	38	339.8	8.942		
PAM	Tissue	1	7.59	7.59	2.63	0.1132
	Residuals	38	109.71	2.89		
SPR	Tissue	1	66.013	66.013	13.069	0.0008683
	Residuals	38	191.943	5.051		

	Df	Pillai	~ F	num Df	den Df	Pr(>F)
Tissue	1	0.32582	5.7994	3	36	0.002429
Residuals	38					

Table 2-4 Multivariate analysis of covariance (MANCOVA) table of genes versus tissue, and Pillai trace.

Supplemental figures, tables, and methods

Tissue	Ontology Type	Ontology Name (GO:ID)	Z Score	Adjusted P
Dorsal	biological process	melanin metabolic process (GO:0006582)	13.77	0.03
Dorsal	biological process	protein heterotrimerization (GO:0070208)	12.23	0.03
Dorsal	biological process	melanocyte differentiation (GO:0030318)	12.02	0.03
Dorsal	cellular component	extracellular matrix (GO:0031012)	12.00	0.03
Dorsal	cellular component	condensin complex (GO:0000796)	11.12	0.03
Dorsal	biological process	extracellular matrix organization (GO:0030198)	10.11	0.03
Dorsal	cellular component	pigment granule (GO:0048770)	9.93	0.03
Dorsal	cellular component	extracellular space (GO:0005615)	9.23	0.03
Dorsal	biological process	cell cycle cytokinesis (GO:0033205)	8.97	0.03
Dorsal	biological process	nuclear division (GO:0000280)	8.87	0.03
Dorsal	biological process	cellular response to radiation (GO:0071478)	8.39	0.03
Dorsal	cellular component	extracellular matrix part (GO:0044420)	8.34	0.03
Dorsal	molecular function	platelet-derived growth factor binding (GO:0048407)	8.05	0.03
Dorsal	biological process	negative regulation of epithelial cell proliferation (GO:0050680)	7.92	0.03
Dorsal	molecular function	identical protein binding (GO:0042802)	7.90	0.03
Dorsal	biological process	neural crest cell migration (GO:0001755)	7.62	0.03
Dorsal	cellular component	cleavage furrow (GO:0032154)	7.35	0.03
Dorsal	biological process	neural crest cell development (GO:0014032)	7.35	0.03
Dorsal	biological process	central nervous system development (GO:0007417)	7.22	0.03
Dorsal	biological process	angiogenesis (GO:0001525)	7.03	0.03
Dorsal	biological process	skeletal system development (GO:0001501)	6.93	0.03
Dorsal	biological process	negative regulation of phosphorylation (GO:0042326)	6.65	0.03
Dorsal	biological process	positive regulation of protein import into nucleus (GO:0042307)	6.46	0.03
Dorsal	biological process	chromosome segregation (GO:0007059)	6.33	0.03
Dorsal	molecular function	glycosaminoglycan binding (GO:0005539)	6.24	0.03
Dorsal	biological process	gonad development (GO:0008406)	5.97	0.03
Dorsal	biological process	hemopoietic progenitor cell differentiation (GO:0002244)	5.93	0.03

Tissue	Ontology Type	Ontology Name (GO:ID)	Z Score	Adjusted P
Dorsal	biological process	cellular process involved in reproduction (GO:0048610)	5.84	0.03
Dorsal	biological process	pigmentation (GO:0043473)	5.75	0.03
Dorsal	biological process	response to acid (GO:0001101)	5.75	0.03
Dorsal	biological process	regionalization (GO:0003002)	5.46	0.03
Dorsal	biological process	regulation of angiogenesis (GO:0045765)	5.17	0.03
Dorsal	biological process	regulation of cell division (GO:0051302)	5.17	0.03
Dorsal	biological process	anatomical structure morphogenesis (GO:0009653)	5.08	0.03
Dorsal	biological process	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway (GO:0090092)	5.06	0.03
Dorsal	biological process	response to retinoic acid (GO:0032526)	5.04	0.03
Dorsal	biological process	regulation of epidermis development (GO:0045682)	5.04	0.03
Dorsal	biological process	regulation of locomotion (GO:0040012)	4.99	0.03
Dorsal	biological process	regulation of cell cycle (GO:0051726)	4.98	0.03
Dorsal	biological process	positive regulation of response to stimulus (GO:0048584)	4.18	0.03
Dorsal	molecular function	protein heterodimerization activity (GO:0046982)	4.09	0.03
Dorsal	biological process	biological adhesion (GO:0022610)	4.03	0.03
Dorsal	biological process	response to ionizing radiation (GO:0010212)	4.01	0.03
Dorsal	biological process	regulation of apoptosis (GO:0042981)	3.94	0.03
Dorsal	biological process	developmental growth (GO:0048589)	3.91	0.03
Dorsal	molecular function	receptor binding (GO:0005102)	3.63	0.03
Dorsal	biological process	regulation of cell adhesion (GO:0030155)	3.62	0.03
Dorsal	biological process	regulation of cell development (GO:0060284)	3.56	0.03
Dorsal	biological process	positive regulation of multicellular organismal process (GO:0051240)	3.46	0.03
Dorsal	biological process	positive regulation of developmental process (GO:0051094)	3.43	0.03
Dorsal	biological process	positive regulation of transcription from RNA polymerase II promoter (GO:0045944)	3.39	0.03
Dorsal	biological process	negative regulation of gene expression (GO:0010629)	3.30	0.03
Dorsal	biological process	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0045934)	3.07	0.03

Tissue	Ontology Type	Ontology Name (GO:ID)	Z Score	Adjusted P
Dewlap	molecular function	ceramidase activity (GO:0017040)	9.77	0.03
Dewlap	biological process	sphingolipid metabolic process (GO:0006665)	9.62	0.03
Dewlap	molecular function	selenium binding (GO:0008430)	8.39	0.03
Dewlap	cellular component	extracellular matrix (GO:0031012)	7.56	0.03
Dewlap	biological process	negative regulation of protein autophosphorylation (GO:0031953)	7.18	0.03
Dewlap	biological process	cellular response to calcium ion (GO:0071277)	7.08	0.03
Dewlap	biological process	lysosome organization (GO:0007040)	7.03	0.03
Dewlap	biological process	response to other organism (GO:0051707)	7.02	0.03
Dewlap	molecular function	bacterial cell surface binding (GO:0051635)	6.65	0.03
Dewlap	molecular function	long-chain fatty acid-CoA ligase activity (GO:0004467)	6.65	0.03
Dewlap	biological process	regulation of type 2 immune response (GO:0002828)	6.57	0.03
Dewlap	biological process	defense response (GO:0006952)	6.51	0.03
Dewlap	cellular component	lysosome (GO:0005764)	6.50	0.03
Dewlap	biological process	negative regulation of carbohydrate metabolic process (GO:0045912)	6.06	0.03
Dewlap	molecular function	heme binding (GO:0020037)	5.98	0.03
Dewlap	biological process	estrogen metabolic process (GO:0008210)	5.85	0.03
Dewlap	cellular component	tropoin complex (GO:0005861)	5.85	0.03
Dewlap	biological process	cellular lipid catabolic process (GO:0044242)	5.77	0.03
Dewlap	molecular function	structural constituent of cytoskeleton (GO:0005200)	5.76	0.03
Dewlap	biological process	proteolysis (GO:0006508)	5.70	0.03
Dewlap	molecular function	peptidase activity (GO:0008233)	5.59	0.03
Dewlap	cellular component	extracellular matrix part (GO:0044420)	5.38	0.03
Dewlap	molecular function	electron carrier activity (GO:0009055)	5.36	0.03
Dewlap	molecular function	oxidoreductase activity (GO:0016491)	5.34	0.03
Dewlap	biological process	regulation of cytokine production (GO:0001817)	5.34	0.03
Dewlap	molecular function	lipase activity (GO:0016298)	5.33	0.03
Dewlap	molecular function	integrin binding (GO:0005178)	5.32	0.03
Dewlap	biological process	immune effector process (GO:0002252)	5.32	0.03
Dewlap	biological process	immune response (GO:0006955)	5.29	0.03
Dewlap	biological process	positive regulation of leukocyte	5.25	0.03

Tissue	Ontology Type	Ontology Name (GO:ID)	Z Score	Adjusted P
		apoptosis (GO:2000108)		
Dewlap	biological process	respiratory burst (GO:0045730)	5.25	0.03
Dewlap	biological process	positive regulation of B cell activation (GO:0050871)	5.23	0.03
Dewlap	biological process	angiogenesis (GO:0001525)	5.10	0.03
Dewlap	molecular function	carbohydrate binding (GO:0030246)	4.97	0.03
Dewlap	biological process	regulation of response to external stimulus (GO:0032101)	4.96	0.03
Dewlap	cellular component	phagocytic vesicle (GO:0045335)	4.78	0.03
Dewlap	biological process	negative regulation of muscle contraction (GO:0045932)	4.70	0.03
Dewlap	biological process	muscle system process (GO:0003012)	4.65	0.03
Dewlap	molecular function	NAD+ ADP-ribosyltransferase activity (GO:0003950)	4.64	0.03
Dewlap	molecular function	morphogen activity (GO:0016015)	4.52	0.03
Dewlap	cellular component	extracellular space (GO:0005615)	4.50	0.03
Dewlap	biological process	cellular response to lipopolysaccharide (GO:0071222)	4.50	0.03
Dewlap	biological process	response to wounding (GO:0009611)	4.47	0.03
Dewlap	biological process	extracellular matrix organization (GO:0030198)	4.42	0.03
Dewlap	biological process	regulation of lymphocyte proliferation (GO:0050670)	4.32	0.03
Dewlap	biological process	regulation of cellular component movement (GO:0051270)	4.31	0.03
Dewlap	biological process	purine ribonucleoside metabolic process (GO:0046128)	4.30	0.03
Dewlap	biological process	Rho protein signal transduction (GO:0007266)	4.30	0.03
Dewlap	biological process	regulation of wound healing (GO:0061041)	4.19	0.03
Dewlap	biological process	negative regulation of lymphocyte activation (GO:0051250)	4.16	0.03
Dewlap	biological process	cell-substrate adhesion (GO:0031589)	4.15	0.03
Dewlap	molecular function	protein phosphatase binding (GO:0019903)	3.97	0.03
Dewlap	biological process	response to extracellular stimulus (GO:0009991)	3.95	0.03
Dewlap	biological process	membrane organization (GO:0061024)	3.90	0.03
Dewlap	molecular function	endopeptidase regulator activity (GO:0061135)	3.78	0.03
Dewlap	biological process	oxidation-reduction process (GO:0055114)	3.39	0.03
Dewlap	cellular component	cell fraction (GO:0000267)	3.39	0.03
Dewlap	biological process	negative regulation of cell death (GO:0060548)	3.37	0.03

Tissue	Ontology Type	Ontology Name (GO:ID)	Z Score	Adjusted P
Dewlap	cellular component	membrane raft (GO:0045121)	3.17	0.03
Dewlap	biological process	negative regulation of cell	2.92	0.03

Supp. Table 2-1. Significant GO categories in significantly expressed gene sets.

Significantly expressed genes are from RNA-seq Cufflinks analysis include all unique dewlap genes significantly expressed relative to the dorsal and ventral skin and all dorsal genes significantly expressed relative to the dewlap and ventral skin. There were no significantly expressed GO terms in the ventral skin relative to either the dewlap or the dorsal skin. Significance was set at 0.05 with an FDR of 0.05.

Type	Gene Name	Full Name	Citations	Ensembl ID	Position
Carotenoid	<i>scarb1</i>	Scavenger receptor class B type I	(Sundvold <i>et al.</i> 2011)	ENSACAG00000014729	LGb:1343010-1368336
Carotenoid	<i>stard5</i>	StAR-related lipid transfer protein 5	(Soccio <i>et al.</i> 2002; Bhosale <i>et al.</i> 2009)	ENSACAG00000011206	GL343894.1:61620-70956
Carotenoid	<i>rbp4</i>	Retinol-binding Protein 4	(Pointer <i>et al.</i> 2011)	ENSACAG00000007043	GL343219.1:1917896-1927414
Carotenoid	<i>plin2</i>	Perilipin-2 (similar to PLIN)	(Londos <i>et al.</i> 1996; Menon 2000)	ENSACAG00000017672	2:62626568-62646439
Carotenoid	<i>cbp/pag1</i>	Carotenoid binding protein homolog	(Walsh <i>et al.</i> 2012)	ENSACAG00000013517	4:24768070-24783457
Carotenoid	<i>gstp1</i>	Glutathione S-transferase alpha 1	(Bhosale & Bernstein 2005)	ENSACAG00000003940	1:90672033-90681467
Carotenoid	<i>gsta2</i>	Glutathione S-transferase alpha 2 (5-prime end)	(Fukamachi <i>et al.</i> 2001; Nakayama <i>et al.</i> 2002; Bhosale & Bernstein 2005)	ENSACAG00000008898	1:151781700-151787309
Carotenoid	<i>cd36</i>	Cluster of differentiation 36	(Sakudoh <i>et al.</i> 2010; 2013)	ENSACAG00000015775	5:93087939-93120933
Carotenoid	<i>bco2</i>	Beta-carotene dioxygenase	(Eriksson <i>et al.</i> 2008)	ENSACAG00000015299	GL343973.1:16622-36304
Carotenoid	<i>bcmol</i>	Beta,beta-carotene 15,15'-monooxygenase	(Kiefer <i>et al.</i> 2001)	ENSACAG00000009055	LGC:5809536-5835833
Iridophore	<i>pnp</i>	Purine nucleoside phosphorylase	(Curran <i>et al.</i> 2010)	ENSACAG00000008033	GL343220.1:1255734-1262614
Iridophore	<i>ltk</i>	Leukocyte Tyrosine Kinase	(Lopes <i>et al.</i> 2008)	ENSACAG00000013890	1:41513957-41585098
Keratin	<i>krtcap2</i>	keratin		ENSACAG00000026489	AAWZ02041356:3061-5431
Keratin	<i>krt80</i>	keratin		ENSACAG00000008392	2:95942478-95980050
Keratin	<i>krt7</i>	keratin		ENSACAG00000007961	2:95868213-95891124

Type	Gene Name	Full Name	Citations	Ensembl ID	Position
Keratin	<i>krt222</i>	keratin		ENSACAG00000016060	6:70363392-70386111
Keratin	<i>krt19</i>	keratin		ENSACAG00000017868	6:70650435-70661158
Keratin	<i>krt18</i>	keratin		ENSACAG00000011762	GL343250.1:1886050-1901529
Keratin	<i>krt12</i>	keratin		ENSACAG00000017866	6:70460683-70467025
Melanin	<i>slc24a5</i>	Cation-exchanger	(Lamason <i>et al.</i> 2005)	ENSACAG00000004764	GL343561.1:312126-321076
Melanin	<i>oca2</i>	Human iris pigmentation	(Frudakis <i>et al.</i> 2003; 2007)	ENSACAG00000011098	3:111316775-111448389
Melanin	<i>aim11</i>	Absent in melanoma 1-like	(Fukamachi <i>et al.</i> 2001; Nakayama <i>et al.</i> 2002)	ENSACAG00000017767	GL343480.1:488078-512682
Melanin	<i>plcz1</i>	Phospholipase C (PLC) pigment dispersion	(Graminski <i>et al.</i> 1993)	ENSACAG00000017592	5:9439657-9471989
Melanin	<i>plch2</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Baynash <i>et al.</i> 1994; Dupin & Le Douarin 2003)	ENSACAG00000011901	GL343848.1:38472-103113
Melanin	<i>plcg2</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Baynash <i>et al.</i> 1994; Dupin & Le Douarin 2003)	ENSACAG00000009200	LGc:6057435-6100308
Melanin	<i>plce1</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Le Guyader <i>et al.</i> 2005)	ENSACAG00000007404	GL343219.1:1336469-1441519
Melanin	<i>plcd3</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler <i>et al.</i> 2000; Braasch <i>et al.</i> 2007)	ENSACAG00000007219	6:64798668-64832366
Melanin	<i>plcb4</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Giordano <i>et al.</i> 2003)	ENSACAG00000002271	1:134864565-134959468
Melanin	<i>plcb3</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000000781	GL343898.1:18766-81386
Melanin	<i>plcb2</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000014083	GL343264.1:1295530-1363849
Melanin	<i>plcb1</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler	ENSACAG00000002652	1:135261540-135440581

Type	Gene Name	Full Name	Citations	Ensembl ID	Position
Melanin	<i>plchl</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000004239	3:15966445-15991014
Melanin	<i>plcg1</i>	Phospholipase C (PLC)	(Graminski <i>et al.</i> 1993; Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000011128	GL343291.1:862841-946174
Melanin	<i>mitf</i>	Microphthalmia-Associated Transcription Factor	(Graw <i>et al.</i> 2003)	ENSACAG00000013586	2:181607271-181633021
Melanin	<i>mc4r</i>	Melanocortin receptor 4	(Takahashi & Kawauchi 2006)	ENSACAG00000007935	GL343408.1:53303-54239
Melanin	<i>mc2r</i>	Melanocortin receptor 2	(Takahashi & Kawauchi 2006)	ENSACAG00000013103	4:43439046-43440787
Melanin	<i>mc1r</i>	Melanocortin receptor 1	(Nachman <i>et al.</i> 2003; Hoekstra & Nachman 2003; Rosenblum <i>et al.</i> 2004)		
Melanin	<i>kit</i>	Tyrosine kinase	(Nakayama <i>et al.</i> 1998; Wilson <i>et al.</i> 2004)	ENSACAG00000009902	5:104401497-104453694
Melanin	<i>foxn1</i>	Ancestral founder mutation of the nude	(Weiner <i>et al.</i> 2007)	ENSACAG00000012187	GL343470.1:488411-497837
Melanin	-	<i>Kit</i> ligand	(Wehrle-Haller <i>et al.</i> 2001)	ENSACAG00000011718	5:33321823-33337662
Melanin	<i>erbb3</i>	EGFR-like tyrosine kinase	(Baynash <i>et al.</i> 1994; Budi <i>et al.</i> 2008)	ENSACAG00000005090	GL343198.1:3672169-3705750
Melanin	<i>cdh11</i>	Cadherin-11	(Greenwood <i>et al.</i> 2012)	ENSACAG00000015087	LGc:8456499-8508931
Melanin	<i>erbb2</i>	Endothelin receptor B subtype 2	(Baynash <i>et al.</i> 1994; Dupin & Le Douarin 2003)	ENSACAG00000015705	GL343202.1:1822582-1837760
Melanin	<i>silv /pmel</i>	Silver	(Hearing & Tsukamoto 1991)	ENSACAG00000005471	GL343198.1:3338817-3353336
Melanin	<i>tyrp1</i>	Dopachrome tautomerase	(Hearing & Tsukamoto 1991; Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000011322	2:33462979-33481653

Type	Gene Name	Full Name	Citations	Ensembl ID	Position
Melanin	<i>tyrp1</i>	Tyrosinase-related protein 1 precursor	(Hearing & Tsukamoto 1991; Curran <i>et al.</i> 2010)	ENSACAG00000011322	2:33462979-33481653
Melanin	<i>tyr</i>	Tyrosinase	(Hearing & Tsukamoto 1991; Giordano <i>et al.</i> 2003)	ENSACAG00000014963	3:199047443-199123973
Melanin	<i>aim11</i>	Absent in melanoma 1-like protein	(Fukamachi <i>et al.</i> 2001; Nakayama <i>et al.</i> 2002)	ENSACAG00000017767	GL343480.1:488078-512682
Melanin	<i>ednrb</i>	Endothelin receptor b1	(Baynash <i>et al.</i> 1994; Dupin & Le Douarin 2003)	ENSACAG00000016202	3:98046708-98070355
Melanin	<i>ednra</i>	Endothelin receptor a	(Baynash <i>et al.</i> 1994; Giordano <i>et al.</i> 2003; Dupin & Le Douarin 2003)	ENSACAG00000003899	5:134135513-134164664
Pteridine	<i>mycbp2</i>	<i>Esrom, mycbp2, pam</i>	(Le Guyader <i>et al.</i> 2005)	ENSACAG00000000646	3:86475748-86635685
Pteridine	<i>xdh</i>	Xanthine dehydrogenase	(Ziegler <i>et al.</i> 2000; Braasch <i>et al.</i> 2007)	ENSACAG00000006868	1:250635453-250698915
Pteridine	<i>txn11</i>	Thioredoxin-like-1 (clot homolog)	(Giordano <i>et al.</i> 2003)	ENSACAG00000009121	GL343213.1:1444610-1458601
Pteridine	<i>spr</i>	Sepiapterin reductase	(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000001020	GL343632.1:40292-42818
Pteridine	<i>qdpr</i>		(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000011209	4:125356239-125370648
Pteridine	<i>pts</i>	6-pyruvoyltetrahydropterin synthase	(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000015279	GL343973.1:631-2594
Pteridine	<i>pcbd2</i>		(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000001400	GL343223.1:354519-393394
Pteridine	<i>gchfr</i>	GTP cyclohydrolase I feedback regulatory protein	(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000003309	GL343264.1:488958-516908

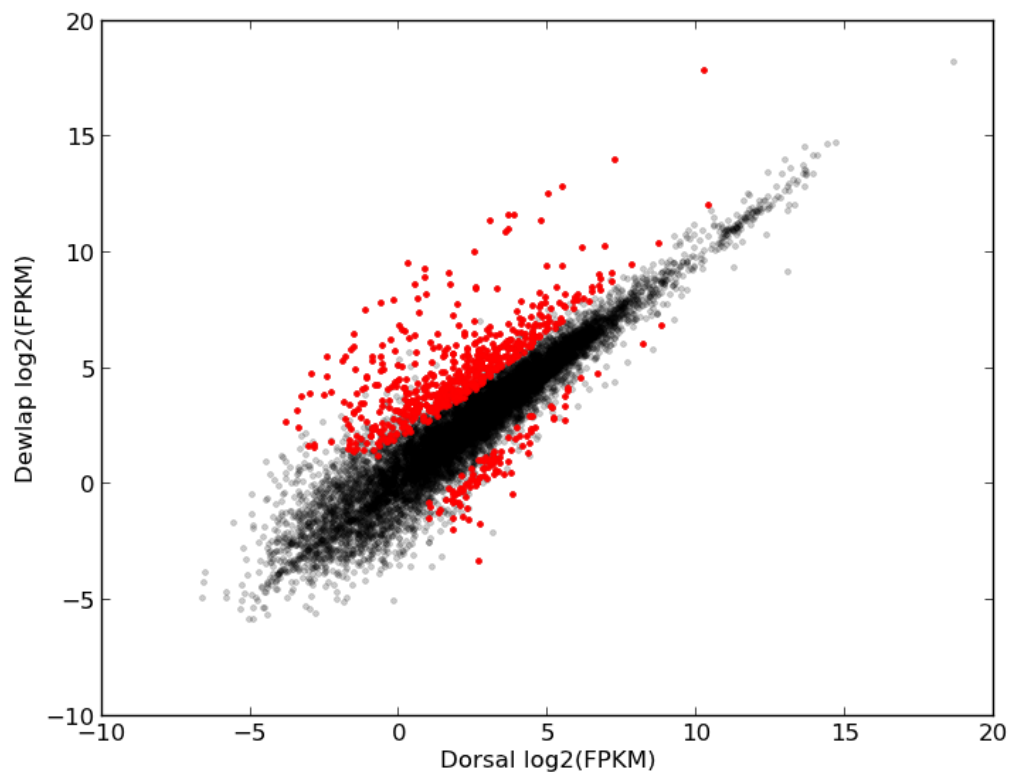
Type	Gene Name	Full Name	Citations	Ensembl ID	Position
Pteridine	<i>gchl</i>	GTP cyclohydrolase I	(Ziegler <i>et al.</i> 2000; Giordano <i>et al.</i> 2003; Braasch <i>et al.</i> 2007)	ENSACAG00000004252	2:142845520-142856658
Pteridine	<i>aox1</i>	Aldehyde oxidase 1 (similar to XDH)	(Curran <i>et al.</i> 2010)	ENSACAG00000016900	1:118008642-118068768
Thioredoxin like	<i>txnrd3</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000014005	2:168601610-168622037
Thioredoxin like	<i>txnrd1</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000015925	5:18317447-18361163
Thioredoxin like	<i>txnl4a</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000001310	4:53139230-53144882
Thioredoxin like	<i>txnl1</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000009121	GL343213.1:1444610-1458601
Thioredoxin like	<i>txnip</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000001167	AAWZ02041877:2100-4231
Thioredoxin like	<i>txndc9</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000008317	GL343456.1:13283-22432
Thioredoxin like	<i>txndc15</i>		(Parichy <i>et al.</i> 2000; Giordano <i>et al.</i> 2003)	ENSACAG00000001427	GL343223.1:332480-343554
Thioredoxin like	<i>txndc11</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000005509	GL343691.1:250368-282035
Thioredoxin like	<i>tmx4</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000003044	1:135860851-135874646
Thioredoxin like	<i>tmx1</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000003935	GL343274.1:211208-220425
Thioredoxin like	<i>prdx2</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000003382	GL343286.1:990298-997204
Thioredoxin like	<i>prdx1</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000015123	4:110190634-110198218

Thioredoxin like	<i>pdia6</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000012985	1:147347755-147361145
Thioredoxin like	<i>erp44</i>		(Giordano <i>et al.</i> 2003)	ENSACAG00000008917	6:55135782-55168669
Xanthopore	<i>csflr</i>	Colony stimulating factor 1 receptor	(Parichy <i>et al.</i> 2000)	ENSACAG00000015531	2:139590627-139613670

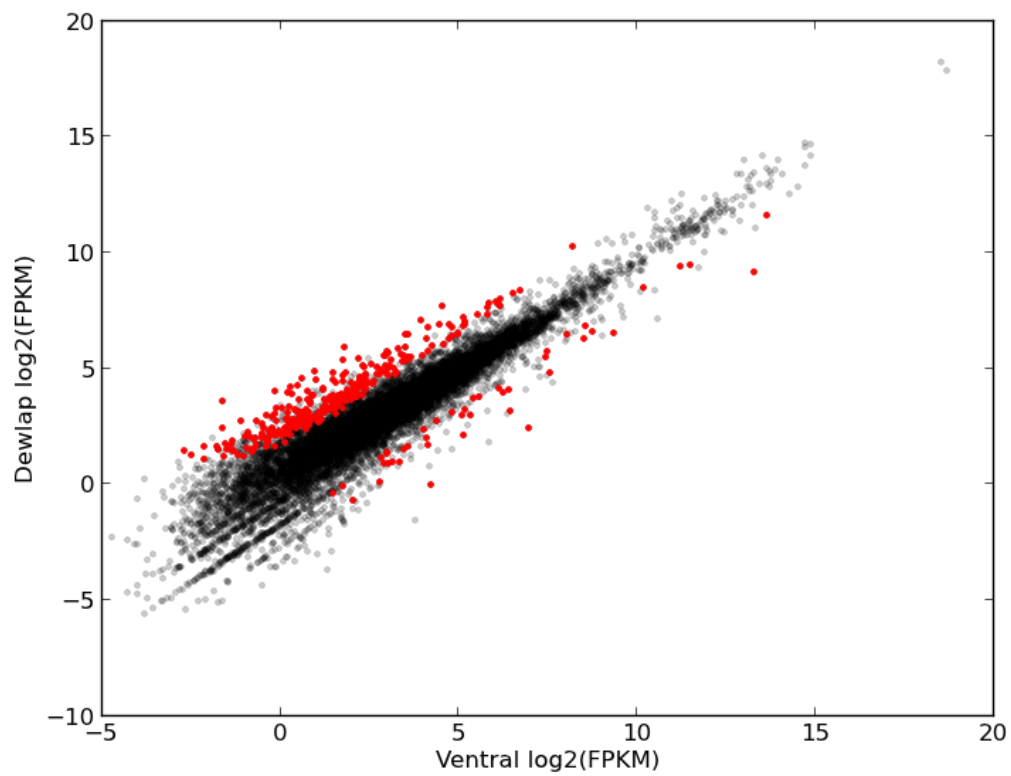
Supp. Table 2-2. Candidate coloration genes identified in this study based on a literature review. Ensembl IDs are from version 67. The position column is formatted as “chromosome:start-stop” and spans both intron and exons.

Gene	Forward Primer	Reverse Primer	Reporter
Beta-Actin (<i>actb</i>)	CGCATTGGCTCCCAGTACAA	CACCAATCCAGACTGAGTATTTGC	CAGGAGGAGCAATGATCTTGA
Protein associated with <i>Myc</i> (<i>pam</i>)	CCCCTATAAACCAGCAAGTGTCA	CACACTGAGCATCAAATTCTATTGTCA	CAGTACAAGGTGACTTTTC
6-pyruvoyl-tetrahydropterin synthase (<i>pts</i>)	CCATCAGCGGCATGTTTCATG	GTCCACATCTTGGTCCAGGTTTT	ATGCAGAAGGCCATCATG
Sepia Pterin Reductase (<i>spr</i>)	CGGACGTGCGTGTTCTCA	CGCACCTCTTCCTGCATGT	ACGCCCCAGGCCCT
Clot homolog	CGACTTGGCAGTAGATGAAGG	ATCGTGCGGAAGGAACTG	Non Applicable

Supp. Table 2-3. Sequences of quantitative PCR primers used in this study.



Supp. Figure 2-1: Log2 fragments per kilobase mapped (FPKM) of dewlap vs dorsal skin. Red points are significant outliers (p -value < 0.05 ; FDR = 0.05).



Supp. Figure 2-2. Log2 fragments per kilobase mapped (FPKM) of dewlap vs ventral skin. Red points are significant outliers (p -value < 0.05 ; FDR = 0.05).

Supp. Method 2-1:

RNA-seq protocol modified from the Cold Spring Harbor Protocol, the Standard Illumina Protocol, the NEB protocol, and the Broad Institute Fisher Paper (Fisher et al. 2011; Zhong et al. 2011).

Extract RNA

1. Remove Tissuelyser insert and chill on ice
2. Add ball bearing and 1000 ul of Trizol to Tissuelyser tube
3. Excise tissue and cut small pieces directly into Tissuelyser tube
4. Homogenize tissue for 2 min at max oscillations (= 50)
5. Cool tube on ice for ~ 1 min
6. Repeat steps 3 and 4
The trizol should appear slightly cloudy at this stage
7. Add 200 ul chloroform (per 1000 ul Trizol)
8. Shake tube vigorously for 15 seconds
9. Incubate tube for 2-3 min at RT
10. Centrifuge for 15 min at 2-8C and 12,000 g
This can be done in the tissuelyzer tube with the ball-bearing present.
11. Carefully pipette and retain the upper phase that contains RNA into an RNase-free 1.5 ml epi tube.
Be careful to avoid the proteinaceous layer between the phases
12. Add 1 ul of glycogen (5 mg/ml)
13. Precipitate RNA with 500ul of isopropyl alcohol
14. Incubate samples for 10 min at RT
15. Centrifuge at 12,000 g for 10 min at 2-8C
16. Remove and discard the supernatant
17. Wash/vortex pellet with at least 1000ul of fresh, cold, 75% EtOH
18. Spin at 7,500g for 5 min at 2-8C
19. Remove supernatant, pipette away extra ethanol
The pellet may be loose so be careful.
20. Dry pellet for 5 min at RT with Kimwipe cover
21. Dissolve by pipetting the RNA pellet in 25 ul RNase free H₂O
22. Incubate for 10 min at 55-60C
23. Cool on ice for 2 min.

Shear RNA

1. Preheat thermocycler to 70C
2. Assemble the following reaction:
 - 1ul 10 x Fragmentation Buffer (Ambion, #AM8740)
 - 10 of total RNA
 - Reserve the remaining tRNA at -80C*
3. Incubate the tube in a PCR thermocycler at 70°C for exactly 5 minutes
4. Add 1 µL of Stop Buffer (Ambion, #AM8740)

5. Put the tube on ice.
6. Transfer contents to 1.5ml centrifuge tube.
7. Assemble the following ingredients
 - 1 μ L of 3M NaOAC pH 5.2
 - 1 μ L of glycogen (5ug/ μ L, Ambion, #AM9510)
 - 30 μ L of 100% EtOH.
8. Incubate the tube at -80°C for 30 minutes.
9. Centrifuge the tube at 12,000g for 25 minutes at 4°C.
10. Remove supernatant taking care not to disturb the pellet
11. Add to the pellet ~ 500 ul of 70% EtOH
12. Vortex and repellet at 7,500g for 5 min.
13. Remove supernatant
14. Air dry the pellet for 5 min
15. Resuspend the RNA in 10 μ L of RNase free water.
16. Transfer to strip tube or capped PCR tube.

First-Strand cDNA Synthesis

1. Assemble the following RT reaction:
 - 10 μ L fragmented mRNA
 - 1.0 μ L Random Hexamer Primers 3ug/ μ L (Invitrogen, #48190-011)
 - 0.5 μ L RNasin Plus or similar
2. Heat at 65 °C for 5 min.
3. Immediately place on ice.
4. Add 7 μ L of the following RT master mix:
 - 4 μ L 5x First Strand Buffer
 - 2 μ L 100mM DTT
 - 1 μ L 10mM dNTPs
5. Mix well.
6. Heat the sample at 25°C in a thermocycler for 2 minutes.
7. Snap chill on ice.
8. Add 1 ul of SuperScript II to the sample.
9. Perform the RT reaction:

Step 1	25°C	10 min
Step 2	42°C	50 min
Step 3	70°C	15 min
Step 4	4 °C	Hold

Second-Strand Synthesis

1. Pre-thaw reagents and chill on ice. Because the rxn is at 16C the reagents must be kept cold.
2. Add the following reagents to the 1st strand mix:
 - 61 μ L of H₂O
 - 10 μ L 10 \times second strand buffer

(500mM Tris-HCl pH7.8, 50mM MgCl₂, 10mM DTT or NEBuffer 2)

3 μ L dNTP mix (10mM)

1 μ L RNaseH (2U/ μ L, Invitrogen, #18021-014)

5 μ L DNA Pol I (10U/ μ L, NEB, #M0209S)

3. Mix well
4. Incubate at 16°C in a thermocycler for 2.5 hours.
 - a. This is a possible stopping point. Eluted dsDNA can be stored at -20°C.
5. SPRI Cleanup with 180 ul of beads.
6. Elute in 20 ul of EB or H₂O. DO NOT remove beads.

End Repair

1. Prepare an appropriate amount of the end-repair master mix on ice as follows:
 - 10.0 μ L NEBNext End Repair Reaction Buffer (10x)
 - 5 μ L End-Repair Mix
 - 85 μ L H₂O
2. Affix lid
3. Incubate at 20°C for 30 min.
4. Remove lid
5. Add 160 ul of 20% PEG 6000, 2.5M NaCl to sample
6. Mix thoroughly
7. Bind on magnet plate
8. Wash with 100 ul of 70% EtOH incubating for 30 sec
9. Remove extra EtOH
10. AirDry Beads

Plus-A

1. Prepare an appropriate amount of the master mix on ice as follows:
 - 5 μ L NEBNext dA-Tailing Reaction Buffer
 - 3 μ L Klenow 3'-5' exo
 - 42 μ L H₂O
2. Affix lid.
3. Incubate at 37°C for 30 min.
4. Remove lid
5. Add 90 ul of 20% PEG 6000, 2.5M NaCl to sample
6. Mix thoroughly
7. Bind on magnet plate
11. Wash with 100 ul of 70% EtOH incubating for 30 sec
12. Remove extra EtOH
13. AirDry Beads

Adapter Ligation

1. Prepare the master mix on ice as follows:
 - 10 μ L NEBNext Quick Ligation Reaction Buffer (5x)

- 5 μ L Quick Ligase
32.5 μ L Di H₂O
2. Add 48 μ L to each well
 3. Add 2.5 μ L of the desired barcode adapter (5 μ M) to each well.
 4. Mix by pipetting up and down.
 5. Affix lid
 6. Incubate at 20C for 15 min.
 7. Remove lid
 8. Add 90 μ L of 20% PEG 6000, 2.5M NaCl to sample
 9. Mix thoroughly
 10. Bind on magnet plate
 11. Wash with 100 μ L of 70% EtOH incubating for 30 sec
 12. Remove extra EtOH
 13. AirDry Beads
 14. Elute in 22 μ L TE (Elution Buffer)
- This is a possible stopping point. The ligated library can be stored at -20°C .*

Library PCR

- 22 μ L of DNA
- 1 μ L of each PCR Primer
- 25 μ L of Kappa Library Amplification MMix

Chapter 3: Genomics of Coloration in *Anolis marmoratus*

Summary

Conspicuous coloration is central to the biology of many animals where it plays a role in social communication, predator avoidance, and mate recognition. In vertebrates, however, almost nothing is known about the genes underlying adaptive conspicuous coloration. To investigate the genomic architecture of conspicuous coloration we examined genomic patterns of divergence among populations of the Guadeloupean anole (*Anolis marmoratus*) at either end of a cline in coloration. One population contains lizards with bright orange spots on their heads and the other contains lizards with blue heads, yet the populations are united by ongoing gene flow. Here we use a low coverage whole genome analysis to compare the complete genomes of 20 individuals, ten from each population. We identify five genomic islands of divergence, which together contain 447 genes, 97 of which contain fixed nucleotide differences between populations. Within the islands of divergence we find two pigmentation genes: melanophilin (*mlph*) and cluster of differentiation 36 (*cd36*). We show that *cd36*, a transmembrane lipoprotein receptor, is up-regulated in the carotenoid containing orange pigmented skin. In animal signaling theory, carotenoid pigments are thought to be honest indicators of quality because they cannot be synthesized endogenously and must be obtained from food. Thus, our results suggest *cd36* may be a key target of sexual selection in anoles.

Introduction

Animal coloration is central to the evolution of many species where it plays pivotal roles in social signaling, predator avoidance and mate recognition (reviewed in: Protas & Patel 2008). Differences in coloration can evolve by both ecological and sexual selection and, because local environments differ across both space and time, ecological and sexual selection can drive divergence in coloration leading to speciation and adaptive radiation (Gavrilets 2002; Yoder *et al.* 2010).

Because of their importance to adaptation and speciation, the genes involved in producing and regulating coloration and pattern are of great interest (Hoekstra 2006). For example, cryptic coloration, which likely evolves primarily by ecological selection, has been frequently associated with genes in the melanin pathway (Nachman *et al.* 2003; Hoekstra & Nachman 2003; Hoekstra *et al.* 2004; Rosenblum *et al.* 2004). However, when color evolves for defense or as a social signal, it is often bright and conspicuous; either providing a warning to predators that the animal is potentially toxic (aposematism) or providing information important in social and sexual interactions. Many conspicuous colorful traits evolve by sexual selection (Hill 1990; Brooks & Endler 2001; Jordan 2008). However, despite the evolutionary importance of sexual selection and the ‘exciting’ nature of colorful pigmentation, little is known about the genes responsible for colorful phenotypes in vertebrates (Walsh *et al.* 2012).

In vertebrates colorful yellow, red, and orange integument is typically colored by two classes of pigments: pteridines and carotenoids. Pteridine pigments are synthesized

within pigment containing cells by the pteridine synthesis pathway (reviewed in: Ziegler 2003; Braasch *et al.* 2007). The pteridine pigments are ubiquitous in the animal kingdom, producing red and yellow pigments in the eyes of flies as well as in the spots and markings of fish, amphibians, and reptiles (Lederer 1940; Dupont 1958; Bagnara 1961; Ortiz & Maldonado 1966; Ziegler 2003) and the pathway has been shown to be conserved between insects and fish (Braasch *et al.* 2007). The pathway consists of about ten genes that regulate the synthesis of differently colored pigments. Like pteridines, carotenoids also contribute to the yellows, oranges, and reds seen in brightly colored spots of fishes, birds, and reptiles (reviewed in: Olson & Owens 1998). However, in vertebrates, carotenoid pigments cannot be produced endogenously - rather they are sequestered from foods. Because of the complicated path carotenoids travel considerably less is known about genes and pathways involved in carotenoid sequestration. However, eleven genes have been associated with changes in carotenoid pigmentation (Walsh *et al.* 2012).

In the context of sexual selection, carotenoids are considered particularly important because they are difficult for organisms to acquire and may therefore be 'honest indicators' of mate quality (Olson & Owens 1998). Empirical evidence in guppies and house finches supports this hypothesis as carotenoid rich markings are correlated with male reproductive success (Kodric-Brown 1985; Hill 1990; Grether *et al.* 2001; Hill *et al.* 2002; Landeen & Badyaev 2012). Recent work has also correlated parasitic infections with levels of circulating carotenoids suggesting that there are trade offs between parasite resistance and carotenoid pigmentation (Baeta *et al.* 2008).

Therefore, genes involved in carotenoid metabolism and sequestration are of particular

interest as they may strongly affect fitness under both ecological and sexual selection.

Lizards in the genus *Anolis* offer a rich system in which to study colorful pigmentation, comprising more than 400 species which display a vast array of colors and patterns (Nicholson *et al.* 2007; Losos 2009). Coloration matters in anoles because they use visual signals in territorial and mating interactions (Jenssen 1977). In addition, two lines of evidence suggest that divergence in coloration is an important component of speciation in the adaptive radiation of anoles. First, the color of the dewlap, the extensible throat fan used primarily by males in territorial and mating interactions, differs between sympatric species and is an important component of the species recognition system (Williams & Rand 1977; Ord & Martins 2006; Vanhooydonck *et al.* 2009). Second, intraspecific geographic variation among populations occurs primarily in color and pattern of adult males, suggesting that sexual selection on coloration may be an important component of speciation in the genus (Underwood & Williams 1959; Lazell 1972; Williams & Rand 1977). Here we examine genomic divergence between two closely related populations of *Anolis* lizards that differ in color and pattern. Because gene flow is high between the populations (Tarvin *et al.* unpublished) only those regions containing loci under strong ecological or sexual selection, such as genes involved in locally adaptive changes in coloration, should show evidence of divergence.

Within *Anolis* lizards, the Guadeloupean anole (*Anolis marmoratus*) shows striking geographic variation in color and pattern, so much so that twelve subspecies have been described based on differences in adult male color and pattern (**Figure 3-1**) (Lazell

1963; Lazell 1972; Schneider 1996). Coloration varies by habitat type; lizards in dry, open habitat are cryptically colored whereas lizards in wetter, more forested habitats are conspicuously colored suggesting that coloration in phenotypically divergent populations of *A. marmoratus* may evolve in response to both ecological and sexual selection (Muñoz *et al.* 2013).

In this study we focus on populations representing two subspecies, *A. m. marmoratus* from Capesterre and *A. m. speciosus* from Pointe-à-Pitre (**Figure 3-1**). These populations inhabit the eastern side of Basse-Terre and the southwestern part of Grande-Terre respectively. *Anolis m. marmoratus* is found in costal rainforest, and males have bright orange spots on their heads. In contrast, *A. m. speciosus* is found in more open habitats, such as open forest or scrub forest, and males lack red spots and have a pale blue wash on their heads (**Figure 3-1**). The colors are thought to be adaptive for effective signaling in different light environments (Endler 1990). The populations are continuously distributed and color variation is clinal, indicating ongoing gene exchange. Previous analysis of mitochondrial DNA variation suggested that the populations were very closely related with nucleotide diversity of mitochondrial *Cytochrome b* less than or equal to 1% (Schneider 1996).

Our analysis focuses primarily on identifying candidate coloration genes under selection in divergent genomic regions. However, we also provide new insights into the genomic architecture of ecological speciation and divergence with gene flow. This study investigates questions about the number, size, and location of outlier regions or genomic islands in order to learn more about the genetic basis of speciation. In addition, it

investigates the types of genes found within genomic islands, whether mutations typically occur in coding sequences or in regulatory regions, and whether these mutations are most often at single nucleotides or whether they consist of larger insertions, deletions, inversions, or translocations. The results of this study identify novel colorful pigmentation genes, elucidate the genomic architecture of local adaptation and divergence with gene flow, and provide insight into the genomics of speciation.

Methods

Sample Collection.

Populations of *A. m. marmoratus* and *A. m. speciosus* were sampled from Capesterre, on the eastern side of Basse-Terre, and near Pointe-à-Pitre on Grande-Terre (**Figure 3-1**). The populations are about 35 kilometers apart. Ten male lizards were collected from each population for a total of 20 individuals. Approximately 2-3 cm of tail was removed from each lizard and preserved in 100% ethanol.

Library Construction and Analysis.

DNA from each tail-tip was extracted with a Qiagen DNeasy Blood and Tissue kit. An Illumina Nextera DNA Sample Preparation kit was used to shear, uniquely tag, and PCR amplify each DNA sample. Post amplification, all 20 libraries were pooled at equal concentrations and run on two lanes on an Illumina Hi-Seq 2000 (101 base-pair paired-end reads).

Following sequencing, reads were demultiplexed with Cassava and contaminating adapters were removed with Scythe (<https://github.com/vsbuffalo/scythe>). Reads were

aligned to the complete *Anolis carolinensis* genome (anoCar vsn. 2.0.67, ensembl) (Alföldi *et al.* 2011) with Stampy (vsn. 1.0.18). Stampy's substitution rate parameter was set to 0.13 to account for the approximately 40 million years of divergence between the *A. carolinensis* and *A. marmoratus* (Losos 2009). To remove potential PCR duplicates that could falsely inflate single nucleotide variant (SNV) confidence, following alignment duplicate reads were flagged with Picard (vsn. 1.81, <http://picard.sourceforge.net>).

We used the Genome Analysis Tool Kit (GATK, v. 2.2-16-g9f648cb) (McKenna *et al.* 2010; Depristo *et al.* 2011) to identify SNVs and short insertions and deletions (indels). GATK uses a Bayesian genotype likelihood model that simultaneously estimates both genotype and allele frequency. The genotypes can then be recalibrated based on a known set of SNVs (truth set) such that only those SNV that have similar properties to the 'true' sites are retained. GATK also includes a tool that corrects alignments errors near indels.

Following the GATK recommended workflow, we began by running GATK's IndelRealigner to correct improper alignments around indels. Then SNVs were identified using permissive settings (the minimum base quality score was set to two) with GATK's UnifiedGenotyper tool. To further refine identification of variants we used the VariantRecalibrator tool using SNVs identified in a RAD sequencing dataset (=19,912 total sites) with SNV quality scores greater than 500 as a "truth set" (McGreevy *et al* In Prep). We considered SNVs to be real if they were part of the set that recapitulated the truth set with a false discovery rate of 1%.

Population Genomic Analysis.

Individual SNVs fixed between populations were identified using custom software (<https://github.com/ngcrawford/pyppgen>). We used ANNOVAR (Wang *et al.* 2010) to annotate SNVs that intersect with genes as well as to identify synonymous and non-synonymous SNVs. Differentiated regions between populations were identified with G''_{ST} (Meirmans & Hedrick 2011) calculated using custom software (<https://github.com/ngcrawford/pyppgen>). Tajima's D was calculated using VCFtools (Danecek *et al.* 2011) in 5 kbp non-overlapping blocks. Only SNVs for which there were five samples per population were used when calculating G''_{ST} and Tajima's D . We choose 5 kbp as our blocks size because this was the minimum blocks size where each blocks contained, on average, enough SNVs ($\bar{x}=109.12$; ± 52.00 stdev) to calculate summary statistics without sacrificing precision.

To measure the extent of linkage disequilibrium along the genome we used Beagle (vsn. 3.3.2) to phase our genotypes (Browning & Browning 2007). Beagle assigns alleles to chromosomes using a hidden markov model. We then used VCFtools to calculate correlation coefficients (r^2) between all SNVs in 25 kbp non-overlapping blocks. Blocks smaller than 25 kbp did not contain enough variants to accurately fit decay curves. Because r^2 is sensitive to rare alleles (Remington *et al.* 2001), we only included SNVs where the minor allele frequency was ≥ 0.2 and where the proportion of missing data was less than 20%. To each window we fitted a decay curve (Weir & Hill 1986; Hill & Weir 1988) and measured the fitted r^2 at the midpoint of the window (Alhaddad *et al.* 2013).

Small insertions and deletions (indels) were identified with GATK's UnifiedGenotyper. We removed indels that fell in the lowest quartile of quality scores. Indels fixed between populations were identified and their location relative to genes was assessed with ANNOVAR (Wang *et al.* 2010). Larger structural variants including indels, inversions, and both inter- and intra-chromosomal translocations were identified with BreakDancer (vsn.1.1) (Chen *et al.* 2009) which identifies structural variants by searching for reads that align discordantly relative to their mates. BreakDancer works best with at least 10x coverage so we ran this analysis on all samples pooled by population. Gene ontology analysis was performed with GO elite (Zambon *et al.* 2012). Final statistical analyses were done with a combination of vcftools (Danecek *et al.* 2011), tabix (Li 2011), BEDtools (Quinlan & Hall 2010), pandas (<http://pandas.pydata.org/>), R statistical software (R Core Team 2012), and custom python code.

Results

Sequencing and Structural variation.

We analyzed two lanes of 101 paired-end reads for a total of 8.6Gb of data. Seventy-four percent of the reads aligned to the *Anolis carolinensis* genome version 2. Mean coverage per sample was 1.4x (SD=0.36). Of the 34 million SNVs that passed variant recalibration only 7.6 million were variable within and among the *marmoratus* populations. The remaining 26.4 were variants due to divergence from the reference genome and were not informative for this study. We identified 90,382 short indels that

intersected 135 exons, whereas larger indels and translocations appeared to be largely correlated with low complexity regions.

Analysis of Outliers.

Instead of using F_{ST} to measure divergence between populations we used G''_{ST} which accounts for multiallelic sites, small sample sizes, and a small number of sampled populations (Meirmans & Hedrick 2011). G''_{ST} is defined as (Figure 4, Meirmans & Hedrick 2011),

$$G''_{ST} = \frac{G'_{ST(Nei)}}{1 - H_S} = \frac{k(H_T - H_S)}{(kH_T - H_S)(1 - H_S)}$$

where H_T is the total gene diversity, H_S is within-population gene diversity, k is the number of sampled populations, and G'_{ST} is from equation 4b in Hedrick (2005). G''_{ST} is similar to G'_{ST} in that it ranges from 0 to 1 when populations have unique sets of alleles, but G''_{ST} is additionally corrected to account for a small number of populations (Meirmans & Hedrick 2011). This means that, when a small number of populations is sampled, values of G''_{ST} will be larger than those reported by F_{ST} or G_{ST} . Globally, mean G''_{ST} does not vary amongst chromosomes ($\bar{x} = 0.166$, $SD = 0.106$) with the exception of micro-chromosome LGb ($\bar{x} = 0.257$, $SD = 0.148$) which has significantly higher values (**Figure 3-2**). We identified the top 1 (N=3345, $\bar{x} = 0.5399$) and 5 percent of non-overlapping blocks (N=16725, $\bar{x} = 0.3569$) (**Figure 3-3**). We defined potentially interesting regions as blocks falling within the top 1% of G''_{ST} outliers.

We also calculated Tajima's D which is defined as:

$$D = \frac{d}{\sqrt{\hat{V}}}$$

where d is the difference between the mean number of polymorphisms between each pair of samples and the total number of polymorphic sites. The denominator ($\sqrt{\hat{V}}$) is the square root of the variance of d (i.e., the standard deviation). Tajima's D can be interpreted both in the context of demography and in the context of selection. When D is positive it indicates a decrease in population size or balancing selection. Conversely, when D is negative it indicates an increase in population size or purifying selection. There is a significant difference in the global mean's of Tajima's D for Capesterre ($\bar{x}=0.770$ $SD=0.636$) and Pointe-à-Pitre ($\bar{x}=0.462$, $SD=0.626$) (t-test; $p \ll 0.001$; **Figure 3-4**). Similarly to G''_{ST} , Tajima's D is significantly higher in both populations on micro-chromosome LGb. In contrast, Tajima's D is significantly lower in the 5 kbp blocks with G''_{ST} in the top 1% in both Capesterre (t-test; $p \ll 0.001$) and Pointe-à-Pitre (t-test; $p = 0.03201$). Analysis of individual SNVs identified 1,465 that are fixed between the two populations. Of these, 19 fall within genes and of those 10 are nonsynonymous (**Supp. Table 3-1a,b**).

Genomic Architecture.

We focused on three main measures of genome architecture: differentiation between populations in 5kbp blocks, which we measured with G''_{ST} , the deviation from equilibrium in the intra-population site frequency spectrum in 5 kbp blocks which we measured with Tajima's D , and linkage disequilibrium (LD) between SNVs which we measured by fitting decay curves to correlated SNVs in 25kb blocks. The most obvious

observation is that both genome wide divergence and LD are low. Mean G''_{ST} is 0.166 (± 0.106 stdev) and mean LD is 0.024 (± 0.042 stdev). The low overall divergence between populations is concordant with microsatellite loci and mitochondrial DNA measured in other pairs of populations on Guadeloupe (Schneider 1996; Muñoz *et al.* 2013). Low values of LD suggest that gene flow between these two populations and recombination are producing a strong homogenizing effect over most of the genome.

Both populations show a slightly positive mean Tajima's D across c. 334,000 windows (**Figure 3-4**). Positive values of Tajima's D indicate population bottlenecks or balancing selection (Fu & Li 1993). One explanation is that populations of *A. marmoratus* and *A. m. speciosus* may have experienced a reduced effective size in their recent history (Schneider 1996).

From a chromosomal perspective mean G''_{ST} does not differ between the six macrochromosomes and seven microchromosomes with the exception of microchromosome LGb which has significantly higher G''_{ST} (**Figure 3-2**). This is similar to the high divergence observed on male sex chromosomes in other pairs of incipient species such as *Ficedula* flycatchers (Ellegren *et al.* 2012). However, in *A. carolinensis* LGb is the female sex chromosome (Alföldi *et al.* 2011). Assuming that LGb is the female sex chromosome in *A. marmoratus*, its divergence could have at least three explanations: (1) Reduced effective population size of the female chromosome is increasing neutral divergence (Charlesworth *et al.* 1987); (2) selection on female genes in LGb is driving divergence (Charlesworth *et al.* 1987); or (3) limited female dispersal

which would result in increased divergence on the female chromosome. Sex biased dispersal has been observed in other lesser Antillean anoles (Johansson *et al.* 2008). Of course male sex chromosomes should diverge the even more rapidly because they do not recombine and are subject to more cell divisions (Graves 2006). Thus, an additional possibility is that LGb may actually be the male sex chromosome in *A. marmoratus*.

When viewed on a genome wide scale, it is apparent that the top 1% of 5 kbp G'_{ST} blocks are primarily clustered into five genomic islands of divergence found on macrochromosomes 1, 2, 3, 5 and 6 with singleton 5 kbp outliers dispersed throughout the genome (**Figure 3-5**). Within the five genomic islands of divergence, Tajima's D is significantly reduced and LD is significantly increased relative to the rest of the genome (**Figure 3-6**). Together this suggests that the five regions of high differentiation are under purifying selection in the two populations.

Within the five divergent islands are 447 genes. These regions contain a number of interesting genes including integument and pigmentation genes (*abca12*, *cd36*, *fox12*, *mlph*, *mocos*, *mreg*), spermatogenesis genes (*asun*, *rsbn11*, *spag16*, *spats2l*), and a thermoregulatory gene (*trpm8*). However, when we examined the distribution of fixed SNVs we observed that they generally clustered at the center of the divergent regions possibly representing population specific haplotypes (**Figure 3-5**). These regions only contain 97 genes (**Supp. Table 3-1a,b**) and of these 97 genes only two, cluster of differentiation 36 (*cd36*) and melanophilin (*mlph*), are associated with pigmentation. Eight SNVs fall within *mlph* introns and five within *cd36* intron. Neither *cd36* nor *mlph*

is intersected by a SNV that makes a change to the coding sequence.

Discussion

In the early stages of speciation, populations are often proximate, both genetically and physically, as organisms locally adapt to new environments. This observation has helped develop a new genome-centric theory of speciation with gene flow (Nosil 2008; Yeaman & Whitlock 2011; Via 2011; Nosil & Feder 2012; Feder *et al.* 2012). The basic principle is summarized as follows: in the earliest phases of speciation selection acts directly on those loci critical to fitness in the new environment. Over time this reduces recombination in these regions producing linkage disequilibrium around the selected loci. Because different loci are selected in the different environments these regions appear as ‘islands of divergence’ in the genome in comparisons between populations evolving in different environments. The size of the ‘islands of divergence’ is determined by the rate of introgression and recombination as well as the strength of selection. The evolution of ‘islands of divergence’ may result in the evolution of reproductive barriers owing to evolved differences in loci directly affecting fitness as well as mutations in the flanking linkage groups. Any degree of reproductive isolation then promotes more rapid divergence across the genome (Feder *et al.* 2011). Empirical evidence is scarce, but ‘islands of divergence’ have been observed in several examples of divergence with gene flow in populations of mosquitos, butterflies, *Rhagoletis* flies, stickleback fish, flycatchers, and house mice (Harr 2006; Neafsey *et al.* 2010; Lawniczak *et al.* 2010; Ellegren *et al.* 2012; Nadeau *et al.* 2012; Jones *et al.* 2012; Michel *et al.*). Together this

suggests that ‘islands of divergence’ may play important roles in the evolution of new species.

The exact nature of how divergent islands evolve is the subject of some debate. One possibility is that loci under strong divergent selection drag along neighboring regions in a process known as divergence hitchhiking (Via 2011). If divergence hitchhiking is playing a significant role, the model suggests that islands should be several megabases in size. Additionally the divergence hitchhiking model posits that selection should be sufficient to produce ‘islands of divergence’ and that structural changes such as inversions or translocations to positions near centromeres are not required. Alternately, it is possible that selection acting on multiple loci can only produce small islands unless the loci occur in regions of low recombination such as inversions or near centromeres (Feder & Nosil 2010). Our results suggest that between populations of *A. marmoratus* genomic islands of divergence are evolving by divergence hitchhiking. Our islands are large, c. 1-2 megabases in size, have significantly increased levels of LD and low values of Tajima’s D. Additionally, they do not appear to be bracketed by inversions and are not particularly close to pericentromeric regions. It seems likely that selection, rather than structural changes of genomic location, is driving the formation of these islands of divergence in populations of *A. marmoratus*.

Our analysis of pigmentation genes within islands identified two likely candidates: melanophilin (*mlph*) and cluster of differentiation 36 (*cd36*). Melanophilin has been shown to affect coloration in cats, dogs, quail, and chickens. Melanophilin mutants have reduced eumelanin and pheomelanin (Ishida *et al.* 2006; Drogemueller *et*

al. 2007; Welle *et al.* 2009; Bed'hom *et al.* 2012) which are characterized by bluish phenotypes in formerly melanized hair or feathers. Mutations in *mlph* result in the defective transport of melanosomes. Although it is tempting to speculate that mutations in *mlph* are contributing to the blue phenotypes in *A. m. speciosus*, blue pigments in anoles are produced by iridophores not melanocytes (Rohrlich & Rubin 1975).

Cd36 is a class B scavenger receptor (*scarb*), a type of transmembrane protein that mediates the uptake of low density lipoproteins (LDLs). Carotenoid pigments are transported as LDLs and *cd36* is associated with the deposition of carotenoid pigments in fish and insects (Kiefer 2002; Sakudoh *et al.* 2010; Sundvold *et al.* 2011; Sakudoh *et al.* 2013). Furthermore, variation in *cd36* has been shown to affect uptake of β -carotene in the silk gland of the silk worm (*Bombyx mori*) (Sakudoh *et al.* 2010; Sakudoh 2013). In the carotenoid-containing dewlap of green anoles (*A. carolinensis*) *cd36* is highly expressed (Chapter 2: this dissertation). This suggests that divergent selection acting on *cd36* is likely involved in the deposition of orange carotenoid pigments in *A. m. marmoratus*.

Conclusion.

The discovery that the carotenoid receptor *cd36* may be under divergent selection suggests that *cd36* is contributing to the orange phenotype in the skin of adult male *A. m. marmoratus*. This is particularly exciting because carotenoid pigments are considered to be honest indicators of quality and play important roles in sexually selected social signals. Despite research on the genetic basis of carotenoid based phenotypes in birds and guppies (Tripathi *et al.* 2009; Walsh *et al.* 2012), no genes have been

previously reported to be involved in sexually selected carotenoid pigmentation. Our results suggest that *cd36* may be a critical gene regulating the deposition of carotenoid pigments in an exclusively male phenotype thus may be an important target of sexual selection.

From a broader perspective, sexual selection and the incidence of sexual dichromatism have been shown to correlate with increased species diversity (Barraclough *et al.* 1995). *Anolis* lizards are remarkably diverse, with more than 400 described species (Losos 2009) and each species possesses a uniquely colored, conspicuous dewlap. Within anoles, differences in coloration and pattern appear to evolve prior to the evolution of larger morphological changes (Losos 2009). This suggests that divergence in pigmentation is an important component of speciation in anoles that sets the stage for ecological diversification of reproductively isolated species. The genomic islands of divergence observed in this study, and the genes contained within them, provide insight into the early stages of speciation within anoles.

Future work in anoline systems similarly characterized by intra-specific variation such as the *Anolis distichus*, *brevirostris*, and *apletohallus* species complexes will help characterize whether the same regions and genes are contributing to early stages of speciation in anoles (Stapley *et al.* 2011; NG & Glor 2011; Lambert *et al.* 2013). And, additional studies of expression and molecular evolution in *mlph* and *cd36* in a broad panel of anoles selected from across the phylogeny will help determine if these genes are important beyond the subspecies of *A. marmoratus*. Of course, our preliminary results showing that *cd36* is differentially expressed in the pink carotenoid containing dewlap of

A. carolinensis already suggests that it may play a larger role in the adaptive radiation of *Anolis* lizards.

Acknowledgements

I thank Martha Muñoz for help organizing the collection of many of the specimens as well as our enthusiastic field assistants Juanita Hopwood, Elbert Mock, and André Schneider. I owe a great deal to Brant Peterson and Harvard research computing for assistance with the Stampy alignments. I also thank Christine Mancuso, Durrell Kapan, the Schneider Lab, and my dissertation committee for comments on early drafts. Last, but not least, I thank the editors and anonymous reviewers whose comments helped to significantly improve this manuscript.

Funding

A Theodore Roosevelt Memorial Grant from the AMNH and NSF Grants DEB-1011544 and DEB-1119734 helped support this research.

Figures and Tables

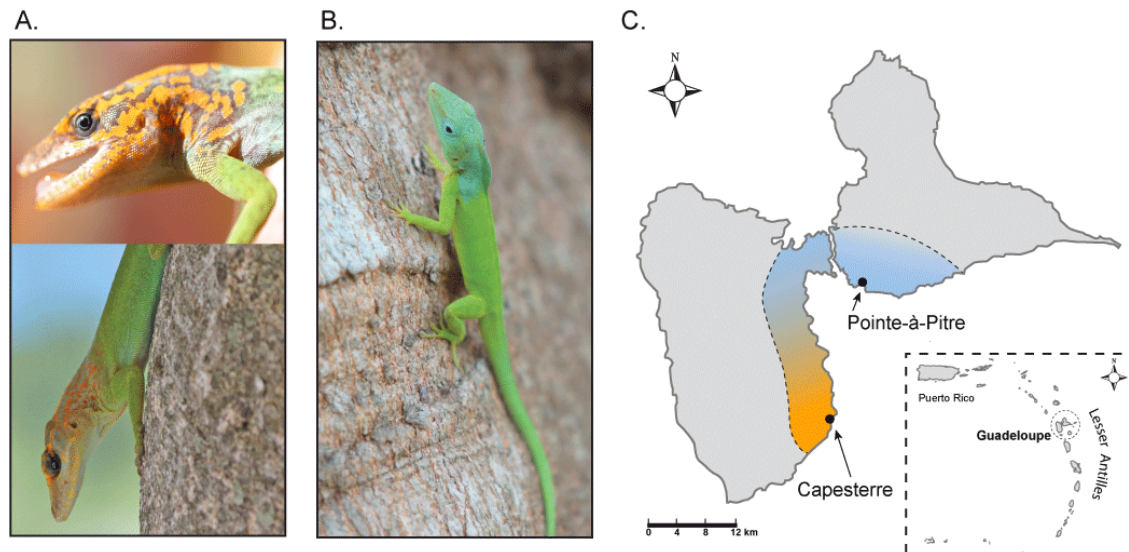


Figure 3-1 **A.** *Anolis marmoratus marmoratus*. **B.** *Anolis marmoratus speciosus*. **C.** Island of Guadeloupe in the French West Indies. Collecting localities are labeled with black circles. The colored region indicates the cline between *A. m. marmoratus* in orange and *A. m. speciosus* in blue. The small inset show position of Guadeloupe within the Lesser Antilles.

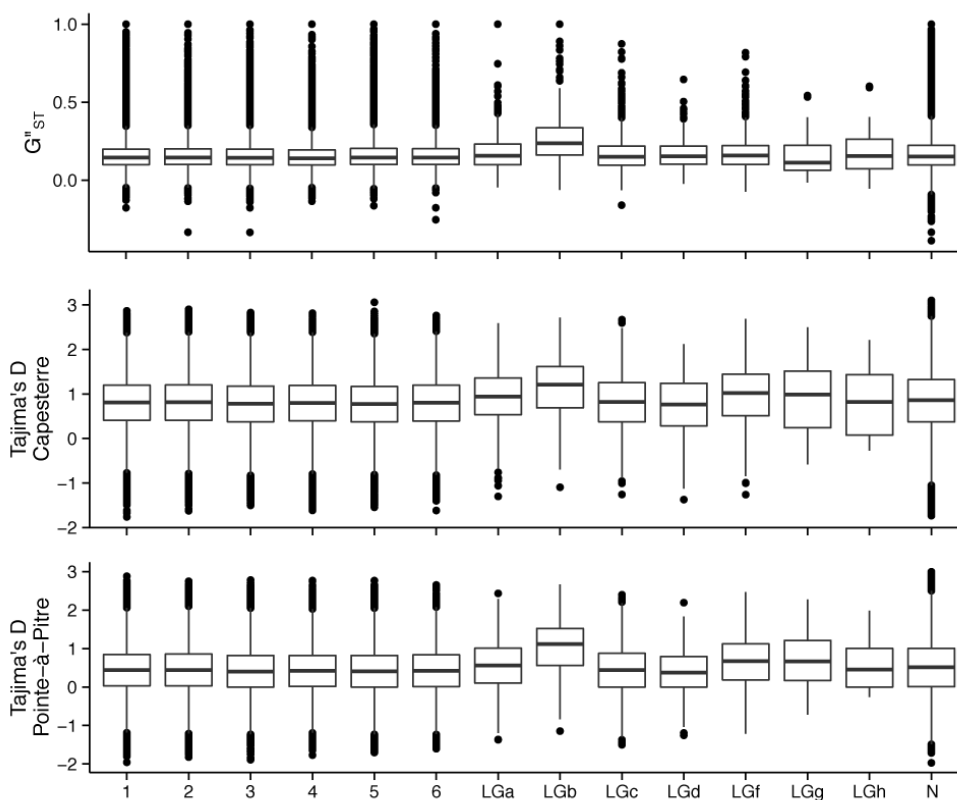


Figure 3-2 Box plots showing mean and variance of G''_{ST} and Tajima's D for each chromosome. Chromosome N represents the combined values for all unassigned scaffolds. LGb shows higher divergence and higher, positive, Tajima's D than other chromosomes and may represent the male sex chromosome.

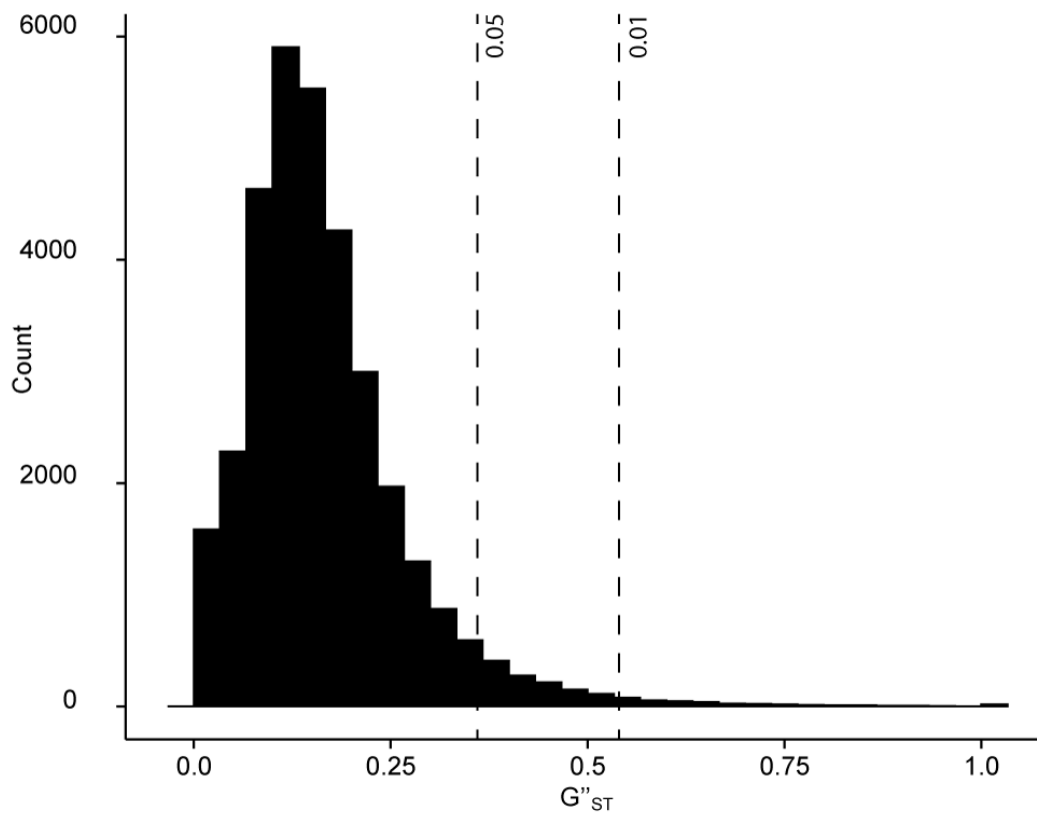


Figure 3-3 Histogram G''_{ST} for all 5 kbp blocks. The dashed lines represent the top 5 and 1 percent of all block respectively.

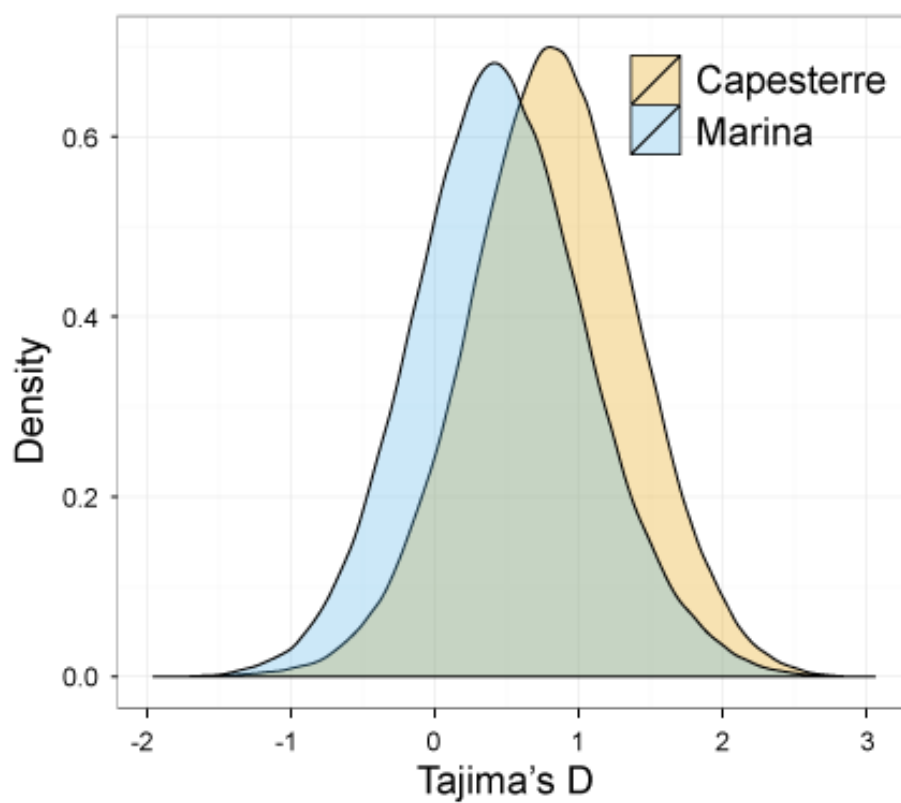


Figure 3-4 Tajima's D from 5 kbp blocks with zero values removed.

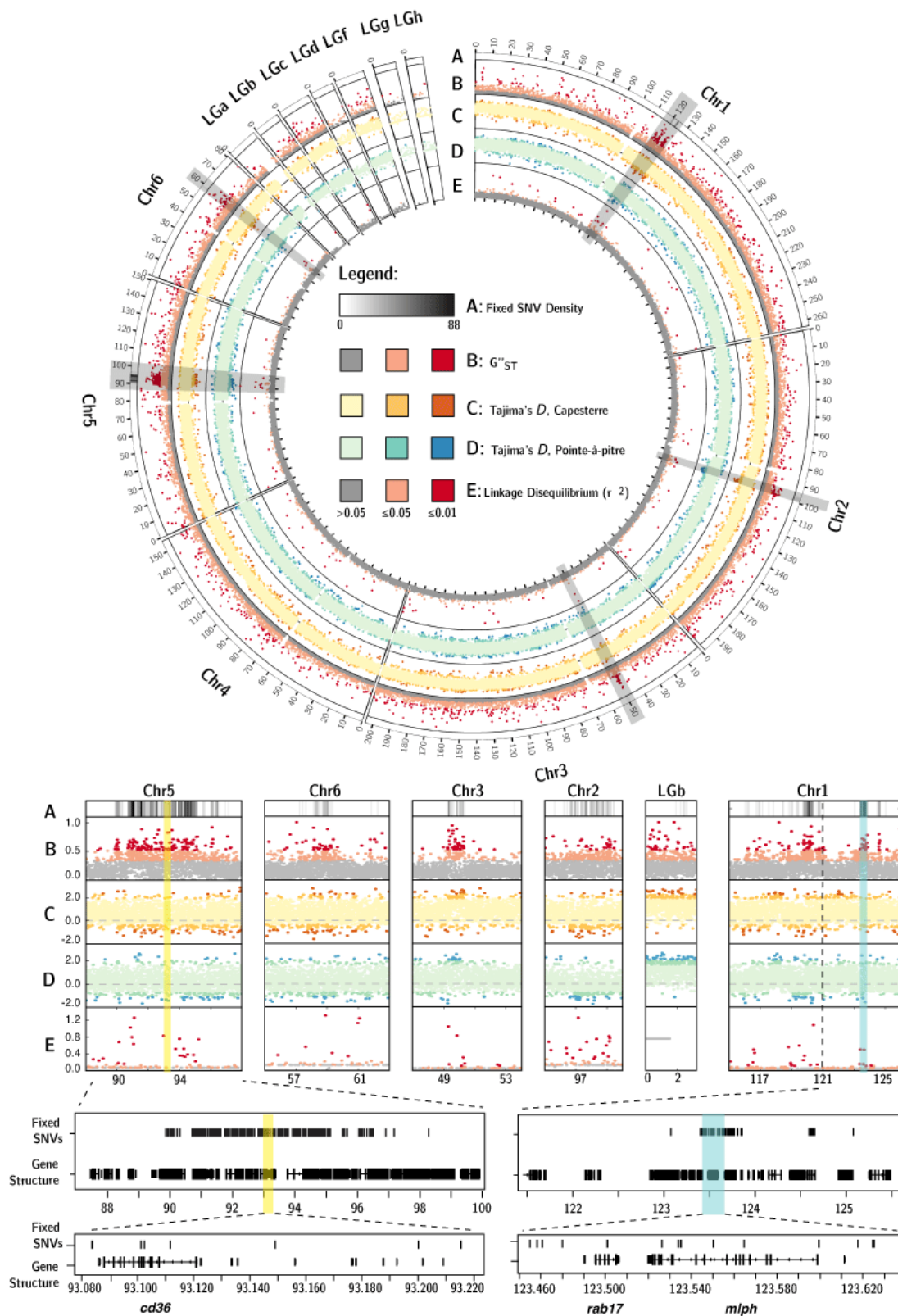


Figure 3-5 The circos plot and the five frames below it display the five summary statistics measured in 5 kbp blocks: **A** = Fixed SNV density (0-88); **B** = G''_{ST} , orange < 0.05, red < 0.01; **C** = Tajima's D Capesterre, light orange < 0.05, dark orange < 0.01, yellow is everything else; **D** = Tajima's D Pointe-à-Pitre, light blue < 0.05, dark blue < 0.01, green is everything else; **E** = median r^2 , orange < 0.05, red < 0.01. The yellow and blue bands indicate the regions where *cd36* and *mlph* are found. The 'zoomed' views in the bottom two rows include the locations of fixed SNVs relative to gene structure. Gene structure is derived from the ASU gene predictions and includes both introns and exons.

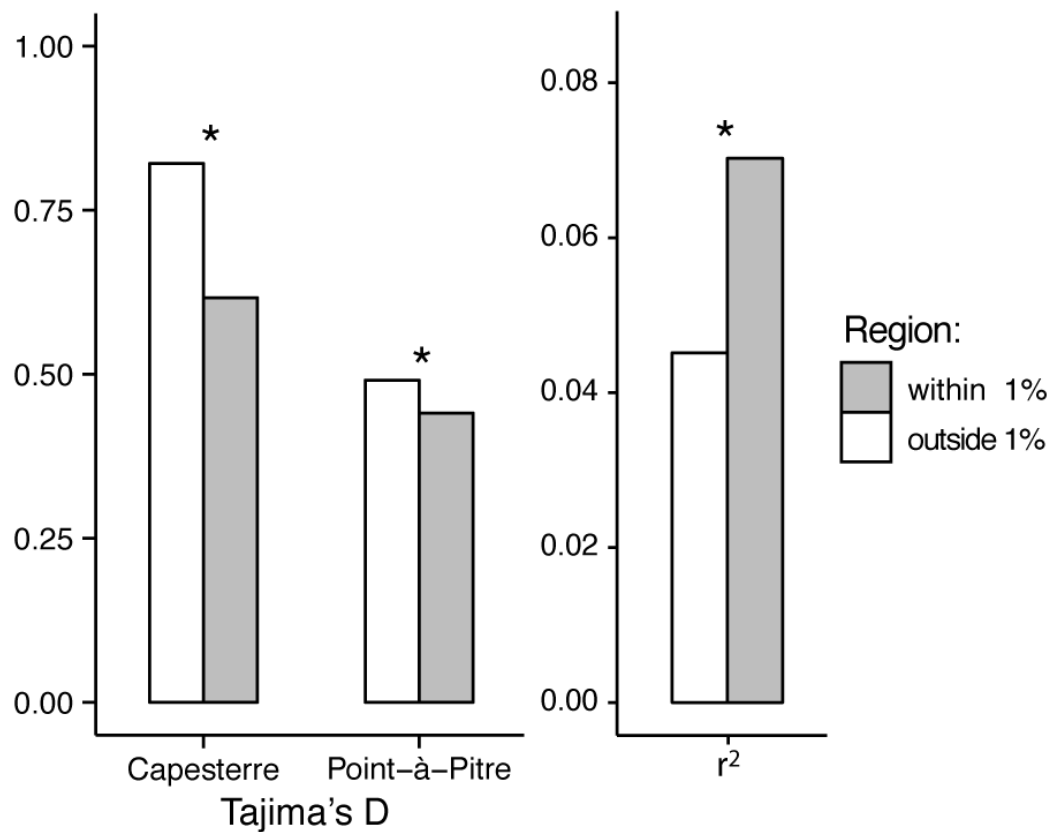


Figure 3-6 Mean G''_{ST} , Tajima's D and R^2 within (grey) and outside of (white) islands of divergence. Students t-test's are highly significant ($p \ll 0.001$) for all comparisons (*) even after Bonferonni correction for multiple comparisons.

Supplemental figures, tables, and methods

Tables:

SNV	Position	Type	Ensembl Gene IDs	Exon Number
A/G	2:98485857	NS	ENSACAG00000015090	3
G/A	2:106120404	NS	ENSACAG00000017630	2
T/C	2:190765588	S	ENSACAG00000010368	16
G/A	4:105651046	NS	ENSACAG00000010960	1
C/T	5:11200617	SG	ENSACAG00000000041	27
G/A	5:89933982	NS	ENSACAG00000001208	14
C/T	5:90948739	NS	ENSACAG000000011690	10
T/A	5:91362527	NS	ENSACAG000000012446	13
A/G	5:91500607	NS	ENSACAG000000012507	9
T/C	5:91502364	S	ENSACAG000000012507	11
T/G	5:92234554	S	ENSACAG000000012663	11
C/T	5:94439413	S	ENSACAG000000009157	4
G/A	5:94762333	S	ENSACAG000000006177	5
T/A	5:94764710	S	ENSACAG000000006177	3
T/A	5:95086045	NS	ENSACAG000000006284	30
T/C	5:95086355	S	ENSACAG000000006284	31
A/G	5:96478795	S	ENSACAG000000006987	23
T/A	6:18082558	NS	ENSACAG000000005361	18
T/G	LGf:2223957	NS	ENSACAG000000003037	4

Supp. Table 3-1a. 19 fixed SNVs with at least 5 samples per population within coding sequences. *Type* abbreviations are: NS = nonsynonymous, S = synonymous, and SG = stopgain. Exon number indicates which exon the SNV is found in.

Gene ID	Gene Name	Expanded Gene Name
ENSACAG00000015090	CCDC137	Coiled-coil domain containing 137
ENSACAG00000017630	Cyp2t4	Cytochrome P450, family 2, subfamily t, polypeptide 4
ENSACAG00000010368	SETD5	SET domain containing 5
ENSACAG00000010960	AMIGO1	Adhesion molecule with Ig-like domain 1
ENSACAG00000000041	OVCH1	Ovochymase 1
ENSACAG00000001208	MCM10	Minichromosome maintenance complex component 10
ENSACAG000000011690	ITPR1	Inositol 1,4,5-trisphosphate receptor, type 1
ENSACAG000000012446	ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2
ENSACAG000000012507	PPFIBP1	PTPRF interacting protein, binding protein 1
ENSACAG000000012663	CACNA2D2	Calcium channel, voltage-dependent, alpha 2/delta subunit 2
ENSACAG000000009157	MAGI2	Membrane associated guanylate kinase
ENSACAG000000006177	PHTF2	Putative homeodomain transcription factor 2
ENSACAG000000006284	PION	Pigeon homolog
ENSACAG000000006987	MLL5	Myeloid/lymphoid or mixed-lineage leukemia 5
ENSACAG000000005361	MRC1	Mannose receptor, C type 1
ENSACAG000000003037	PLD3	Phospholipase D family, member 3

Supp. Table 3-1b. Ensembl Gene IDs to Gene Names.

Chrm	Start	Stop	Name	Fixed SNVs	SNVs/kbp
1	119,835,491	119,838,333	<i>gbx2</i>	2	0.704
1	120,144,930	120,162,167	<i>cxcr7</i>	4	0.232
1	120,295,312	120,333,442	<i>cops8</i>	8	0.210
1	123,843,123	123,856,597	ASU_ACAR_G.861	2	0.148
1	123,674,066	123,702,692	ASU_ACAR_G.516	4	0.140
1	123,667,664	123,792,002	<i>col6a3</i>	17	0.137
1	119,992,527	120,054,757	ASU_ACAR_G.1595	8	0.129
1	124,569,780	124,659,146	<i>spred2</i>	9	0.101
1	123,366,164	123,419,914	<i>pcbp3</i>	5	0.093
1	123,510,012	123,601,715	<i>mlph</i>	8	0.087
1	119,878,878	119,923,750	<i>asb18</i>	3	0.067
1	119,945,098	119,975,852	ASU_ACAR_G.1896	2	0.065
1	123,480,220	123,496,446	<i>rab17</i>	1	0.062
1	120,646,948	120,860,206	<i>wdpcp</i>	7	0.033
1	123,059,468	123,091,621	<i>lss</i>	1	0.031
1	118,795,851	118,843,816	<i>usp40</i>	1	0.021
1	117,147,355	117,196,556	<i>mreg</i>	1	0.020
1	118,939,032	119,001,512	<i>sh3bp4</i>	1	0.016
1	119,299,304	119,366,110	<i>trpm8</i>	1	0.015
1	116,842,704	116,943,930	<i>fn1</i>	1	0.010
1	120,920,940	121,218,361	<i>ehbp1</i>	2	0.007
1	117,398,453	117,547,709	<i>lrrfip1</i>	1	0.007
1	119,494,704	119,801,005	ASU_ACAR_G.1239	2	0.007
1	118,407,016	118,568,515	<i>hdac4</i>	1	0.006
1	115,654,185	116,352,180	<i>spag16</i>	4	0.006
2	98,508,848	98,515,985	ASU_ACAR_G.4239	1	0.140
2	98,774,658	98,803,896	ASU_ACAR_G.2804	4	0.137
2	98,479,182	98,489,550	<i>ccdc137</i>	1	0.096
2	98,286,852	98,348,971	<i>kif19</i>	5	0.080
2	98,411,171	98,445,852	<i>gprc5c</i>	2	0.058
2	98,089,216	98,106,933	<i>rpl38</i>	1	0.056
2	98,489,332	98,508,518	<i>c17orf90</i>	1	0.052
2	98,938,873	98,965,443	ASU_ACAR_G.3616	1	0.038
2	99,066,280	99,174,628	<i>bahcc1</i>	4	0.037
2	98,117,434	98,243,345	<i>ttyh2</i>	3	0.024
2	99,204,345	99,313,730	<i>slc38a10</i>	2	0.018
3	49,341,212	49,372,986	ASU_ACAR_G.6170	5	0.157
3	49,373,287	49,386,266	ASU_ACAR_G.5477	1	0.077

Chrm	Start	Stop	Name	Fixed SNVs	SNVs/kbp
3	49,280,721	49,305,751	ASU_ACAR_G.5128	1	0.040
3	49,386,391	49,454,228	ASU_ACAR_G.5200	2	0.029
3	53,442,663	53,571,381	<i>nps</i>	1	0.008
5	91,047,732	91,061,160	ASU_ACAR_G.8643	4	0.298
5	91,349,266	91,370,664	<i>arntl2</i>	5	0.234
5	94,759,018	94,813,911	<i>phtf2</i>	9	0.164
5	94,734,446	94,753,219	<i>phtf2</i>	3	0.160
5	90,723,760	90,737,227	<i>sspn</i>	2	0.149
5	94,301,911	94,387,054	ASU_ACAR_G.9356	12	0.141
5	91,101,471	91,139,803	<i>tm7sf3</i>	5	0.130
5	90,033,023	90,056,227	<i>sephs1</i>	3	0.129
5	92,565,814	92,574,008	ASU_ACAR_G.9238	1	0.122
5	90,759,970	91,032,630	ASU_ACAR_G.9639	31	0.114
5	92,483,966	92,565,424	ASU_ACAR_G.8729	9	0.110
5	92,481,254	92,547,892	<i>hgf</i>	7	0.105
5	93,356,786	93,387,109	<i>gnai1</i>	3	0.099
5	91,064,533	91,085,807	<i>asun</i>	2	0.094
5	90,076,684	90,152,191	<i>bend7</i>	7	0.093
5	91,391,137	91,542,974	<i>ppfibp1</i>	14	0.092
5	93,761,912	94,694,509	<i>magi2</i>	77	0.083
5	94,882,679	94,967,846	ASU_ACAR_G.9343	7	0.082
5	95,018,720	95,087,720	<i>pion</i>	5	0.072
5	91,261,050	91,318,784	<i>stk38l</i>	4	0.069
5	95,098,726	95,161,797	<i>ccdc146</i>	4	0.063
5	91,546,567	91,578,856	<i>cyb5r3</i>	2	0.062
5	89,911,937	89,944,937	<i>mcm10</i>	2	0.061
5	91,210,091	91,227,438	ASU_ACAR_G.9116	1	0.058
5	96,404,072	96,484,222	<i>mll5</i>	4	0.050
5	91,802,251	92,347,805	ASU_ACAR_G.8645	27	0.049
5	96,338,532	96,402,302	ASU_ACAR_G.9435	3	0.047
5	96,911,503	96,932,942	ASU_ACAR_G.9674	1	0.047
5	93,086,140	93,209,248	<i>cd36</i>	5	0.041
5	92,782,560	93,057,082	<i>sema3c</i>	11	0.040
5	89,976,817	90,029,508	<i>phyh</i>	2	0.038
5	91,143,278	91,170,041	<i>med21</i>	1	0.037
5	96,070,870	96,361,913	<i>lhfp13</i>	9	0.031
5	92,422,593	92,459,019	ASU_ACAR_G.9652	1	0.027
5	93,222,769	93,295,736	<i>gnat3</i>	2	0.027

Chrm	Start	Stop	Name	Fixed SNVs	SNVs/kbp
5	92,609,151	92,719,185	ASU_ACAR_G.9577	3	0.027
5	94,842,790	94,881,580	<i>rsbn11</i>	1	0.026
5	91,145,587	91,185,218	ASU_ACAR_G.9036	1	0.025
5	95,484,817	95,699,674	<i>reln</i>	4	0.019
5	96,483,250	96,616,399	<i>srpk2</i>	2	0.015
5	90,177,605	90,467,331	ASU_ACAR_G.9392	3	0.010
5	97,160,785	97,259,958	<i>prkar2b</i>	1	0.010
6	58,874,533	58,911,992	<i>hgsnat</i>	4	0.107
6	59,047,475	59,095,011	<i>usol</i>	5	0.105
6	58,911,486	58,940,401	<i>ints10</i>	3	0.104
6	59,363,344	59,380,599	<i>fabp1</i>	1	0.058
6	58,293,349	58,385,173	ASU_ACAR_G.1010 8	5	0.054
6	58,753,098	58,847,042	<i>slc20a2</i>	4	0.043
6	58,235,094	58,287,278	<i>uba6</i>	2	0.038
6	59,145,786	59,171,970	<i>cdk12</i>	1	0.038
6	58,395,777	58,471,894	ASU_ACAR_G.1061 9	2	0.026
6	58,520,646	58,563,241	<i>ythdc1</i>	1	0.023
6	61,824,974	61,884,161	<i>ccr10</i>	1	0.017
6	57,422,324	57,490,624	<i>tmem245</i>	1	0.015
6	62,189,732	62,264,752	<i>wnk4</i>	1	0.013
6	57,222,624	57,383,851	ASU_ACAR_G.9876	1	0.006

Supp. Table 3-2 Genes, including introns, within genomic islands of divergence containing at least one fixed SNV.

Appendix

SMOGD: Software for the Measurement of Genetic Diversity

Summary

SMOGD is a web-based application for the calculation of the recently proposed genetic diversity indices G'_{ST} and D_{est} . SMOGD includes bootstrapping functionality for estimating the variance, standard error, and confidence intervals of estimated parameters, and SMOGD also generates genetic distance matrices from pairwise comparisons between populations. SMOGD accepts standard, multilocus Genepop and Arlequin formatted input files and produces HTML and tab-delimited output. This allows easy data submission, quick visualization, and rapid import of results into spreadsheet or database programs.

Main Text

Recently, two diversity measures, D_{est} (Jost 2008) and G'_{ST} (Hedrick 2005), were reported in the literature. These measures more accurately account for differences in allelic diversity than traditional measures such as F_{ST} (Wright 1951; Wright 1965) and G_{ST} (Nei 1973; Nei & Chesser 1983) (reviewed in: Heller & Siegismund 2009; Jost 2009; Ryman & Leimar 2009) especially for highly polymorphic markers such as microsatellite DNA loci. D_{est} further improves on G'_{ST} and G_{ST} , as both incorrectly report that a population is entirely differentiated at a locus when one sub-population is fixed at one allele, but all others are fixed at a different allele (Gregorious et al. 2007). Although many researchers will be interested in calculating D_{est} , and also G'_{ST} , it is not

easy to calculate these measures from typical data sets. For example, Hedrick's (2005) standardized measure of diversity (G'_{ST}) can be calculated with GenoDive (Meirmans 2004). Similarly, Jost's (2008) D_{est} can be calculated with SPADE (Chao & Shen 2008), but by only a single locus at a time. However, GenoDive and SPADE are specific for different computing platforms, and neither calculates both diversity measures simultaneously. Additionally, both GenoDive and SPADE require a software download and installation.

SMOGD calculates G_{ST_est} , G'_{ST} and D_{est} for each locus in a data set and reports the intermediate values (i.e., n , \tilde{N} , H_{S_est} , H_{T_est}) used to calculate the diversity measures. Following the method of Chao (2008) SMOGD can generate up to a thousand bootstrap replicates to calculate variance, standard error, and 95% confidence interval of each diversity measure. SMOGD also computes tables of pairwise comparisons between populations. Lastly, SMOGD provides a default data set containing allele frequencies similar to the first two examples given in Jost (2008, Table 1) for testing purposes.

SMOGD accepts Genepop and Arlequin formatted files (Raymond & Rousset 1995; Excoffier et al. 2005) and runs on any computer with an Internet connection. Results are available as html and tab-delimited files suitable for import into spreadsheet or database programs. A user manual is also available for download.

SMOGD is written in Python 2.5 (Python Software Foundation 2007). It employs the Numpy and Scipy modules that provide fast matrix algebra and statistical methods (Oliphant 2007). Django, a web framework also written in Python, is used to manage the

web interface and back end code (Django Software Foundation 2009). SMOGD may be accessed at: <http://people.bu.edu/ngcrawfo/smogd>. The source code is available from the author upon request.

Pypgen: Calculating a Diversity of Fixation Indexes Across Genomes

Summary

Pypgen is a python package and set of scripts for calculating multi-locus diversity estimators from large population genomic data. Pypgen can calculate G_{ST} as well as G'_{ST} , G''_{ST} , and Jost's D at both the level of individual SNV s as well as across user defined windows. Pypgen is multiprocessed and operates on compressed files.

Availability: The stable release is available on the Python Package Index site from where it can be automatically downloaded and installed. (<http://pypi.python.org/pypi/pypgen>).

The development version is available on github (<https://github.com/ngcrawford/pypgen>)

Introduction

There are many ways to investigate genetic differences between populations, but the classic approach involves summarizing differences in allele frequencies with a fixation index (Wright 1949). Fixation indexes, also known as F -statistics, are diversity measures that range from zero, when populations are undifferentiated, to one when populations have different sets of alleles at a locus. When markers are biallelic, and a large number of samples are present, the classic fixation index F_{ST} is sufficient to describe the differences between populations. However, if fewer samples are present and

the loci are multiallelic then it is necessary to correct for small sample sizes and to incorporate the additional alleles into the statistic. (Nei & Chesser 1983) G_{ST} was developed for this purpose. However, numerous authors have noted that the maximum value of G_{ST} is a function of the intra-population diversity (Hedrick 2005; Jost 2008; Meirmans & Hedrick 2011). This means that G_{ST} 's maximum value is not always equal to one. Furthermore, when populations are fixed for different sets of alleles, G_{ST} underestimates the diversity because the identities of the alleles are lost when they are summarized as intra-population diversity (Jost 2008). To circumvent these limitations a number of corrected diversity estimators have been proposed such as G'_{ST} , G''_{ST} , and Jost's D (reviewed in (Meirmans & Hedrick 2011)).

These measures are of interest to population genomicists for two primary reasons. First, software such as the Genome Analysis Toolkit (GATK) can call multiple alleles at single nucleotide variant (SNV). Although SNVs can only consist of at most four allelic states (i.e., the four nucleotides that make up DNA: A, T, C and G), it is possible to observe a SNV were, for example, one population contains only A and C and the other only the G and T nucleotides. If the intra-population diversity is similar, G_{ST} will find that there is no variation at this SNV. Second, in closely related populations, similar patterns may be observed when different haplotypes are fixed in different populations as SNVs along these haplotypes may be consistently fixed at different nucleotide states. In this scenario windowed measures of G_{ST} may underestimate diversity between populations. This systematic error inherent in the calculation of G_{ST} is important to correct for because

measuring fixation indices along genomes is a first pass approach for identifying outlier SNVs and genomic regions.

Pypgen

To address this problem I wrote pypgen, a python package that contains modules and scripts for calculating G_{ST} , G'_{ST} , G''_{ST} , and Jost's D both at individual SNVs as well as along windows. For the windowed analysis pypgen estimates the multi-locus versions of these measures as well as the standard deviation of each parameter across all SNVs in a particular window.

Because VCF files can contain millions of SNVs, pypgen operates on bgzipped compressed VCF files. In addition to conserving hard-drive space this format allows for fast random access of genomic regions. To enable this functionality, pypgen requires pysam a python module that wraps samtools' tabix interface (<http://www.cgat.org/~andreas/documentation/pysam/>)(Li *et al.* 2009). Pypgen also uses python's multiprocessing module to parallelize the calculations of F -statistics. For the per SNV calculations python's multiprocessing 'imap' method manages portioning the SNVs across processor cores. However, for the windowed analyses pypgen first parses the chromosome names and sizes from the VCF headers to calculate the positions of windows along each chromosome. Then the appropriate regions are extracted using pysam's tabix interface. Next, the regions are mapped across the available processor cores and F -statistics and additional metrics are calculated for each region. The results are then collected and appropriately formatted.

Pyngen is available on the python package index (PyPI) and can be installed on a user's computer with a single command. Pyngen documentation is distributed with the package as well being made available online. Pyngen includes UnitTests for its statistical and VCF parsing methods. Pyngen emits either a comma separated or tab delimited output that is similar to Browser Extensible Data format (BED) (<https://genome.ucsc.edu/FAQ/FAQformat.html#format1>).

Bibliography

- Aitman TJ, Cooper LD, Norsworthy PJ *et al.* (2000) Population genetics: Malaria susceptibility and CD36 mutation. *Nature*, **405**, 1015–1016.
- Alexander NJ, Fahrenbach WH (1969) The dermal chromatophores of *Anolis carolinensis* (reptilia, Iguanidae). *American Journal of Anatomy*, **126**, 41–55.
- Alföldi J, di Palma F, Grabherr M *et al.* (2011) The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, **477**, 587–591.
- Alhaddad H, Khan R, Grahn RA *et al.* (2013) Extent of Linkage Disequilibrium in the Domestic Cat, *Felis silvestris catus*, and Its Breeds. : *Public Library of Science ONE*, **8**, e53537.
- Alibardi L, Segalla A, Dalla Valle L (2012) Distribution of Specific Keratin-Associated Beta-Proteins (Beta-Keratins) in the Epidermis of the Lizard *Anolis carolinensis* Helps to Clarify the Process of Cornification in Lepidosaurians. *The Journal of experimental zoology*, **318**, 388–403.
- Baeta R, Faivre B, Motreuil S, Gaillard M, Moreau J (2008) Carotenoid trade-off between parasitic resistance and sexual display: an experimental study in the blackbird (*Turdus merula*). *Proceedings Of The Royal Society B-Biological Sciences*, **275**, 427–434.
- Bagnara JT (1961) Chromatotrophic hormone, pteridines, and amphibian pigmentation. *General and comparative endocrinology*, **1**, 124–133.
- Bagnara JT, Hadley ME (1973) *Chromatophores and color change*. Prentice Hall.
- Bandarchi B, Ma L, Navab R, Seth A, Rasty G (2010) From Melanocyte to Metastatic Malignant Melanoma. *Dermatology Research and Practice*, **2010**, 1–8.
- Barraclough TG, Harvey PH, Nee S (1995) Sexual Selection and Taxonomic Diversity in Passerine Birds. *Proceedings Of The Royal Society B-Biological Sciences*, **259**, 211–215.
- Baynash AG, Hosoda K, Giaid A *et al.* (1994) Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. *Cell*, **79**, 1277–1285.
- Bed'hom B, Vaez M, Coville J-L *et al.* (2012) The lavender plumage colour in Japanese quail is associated with a complex mutation in the region of MLPH that is related to differences in growth, feed consumption and body temperature. *BioMed Central Genomics*, **13**, 442.

- Bejerano G, Pheasant M, Makunin I *et al.* (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
- Ben J, Lim TM, Phang VPE, Chan W-K (2003) Cloning and Tissue Expression of 6-Pyruvoyl Tetrahydropterin Synthase and Xanthine Dehydrogenase from *Poecilia reticulata*. *Marine Biotechnology*, **5**, 568–578.
- Bhosale P, Bernstein PS (2005) Synergistic effects of zeaxanthin and its binding protein in the prevention of lipid membrane oxidation. *Biochimica et biophysica acta*, **1740**, 116–121.
- Bhosale P, Li B, Sharifzadeh M *et al.* (2009) Purification and partial characterization of a lutein-binding protein from human retina. *Biochemistry*, **48**, 4798–4807.
- Bond AB (2007) The Evolution of Color Polymorphism: Crypticity, Searching Images, and Apostatic Selection. *Annu. Rev. Ecol. Evol. Syst.*, **38**, 489–514.
- Braasch I, Schartl M, Volff J-N (2007) Evolution of pigment synthesis pathways by gene and genome duplication in fish. *BioMed Central Evolutionary Biology*, **7**, 74.
- Bradley BJ, Mundy NI (2008) The primate palette: the evolution of primate coloration. *Evolutionary Anthropology: Issues, News, and Reviews*, **17**, 97–111.
- Brooks R, Endler JA (2001) Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). *Evolution*, **55**, 1002–1015.
- Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics*, **81**, 1084–1097.
- Budi EH, Patterson LB, Parichy DM (2008) Embryonic requirements for ErbB signaling in neural crest development and adult pigment pattern formation. *Development*, **135**, 2603–2614.
- Campbell-Staton SC, Goodman RM, Backström N *et al.* (2012) Out of Florida: mtDNA reveals patterns of migration and Pleistocene range expansion of the Green Anole lizard (*Anolis carolinensis*). *Ecology and Evolution*, **2**, 2274–2284.
- Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M (2000) Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene*, **259**, 139–148.
- Charlesworth B, Coyne JA, Barton NH (1987) The Relative Rates of Evolution of Sex-Chromosomes and Autosomes. *The American Naturalist*, **130**, 113–146.
- Chen K, Wallis JW, McLellan MD *et al.* (2009) BreakDancer: an algorithm for high-

- resolution mapping of genomic structural variation. *Nature Methods*, **6**, 677–681.
- Clotfelter ED, Ardia DR, McGraw KJ (2007) Red fish, blue fish: trade-offs between pigmentation and immunity in *Betta splendens*. *Behavioral Ecology*, **18**, 1139–1145.
- Craig DW, Pearson JV, Szelinger S *et al.* (2008) Identification of genetic variants using bar-coded multiplexed sequencing. *Nature Methods*, **5**, 887–893.
- Curran K, Lister JA, Kunkel GR *et al.* (2010) Interplay between *Foxd3* and *Mitf* regulates cell fate plasticity in the zebrafish neural crest. *Developmental biology*, **344**, 107–118.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- Depristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.
- Dorsky RI, Moon RT, Raible DW (1998) Control of neural crest cell fate by the Wnt signalling pathway. *Nature*, **396**, 370–373.
- Drogemueller C, Philipp U, Haase B, Guezel-Apel A-R, Leeb T (2007) A Noncoding melanophilin gene (MLPH) SNP at the splice donor of Exon 1 represents a candidate causal mutation for coat color dilution in dogs. *Journal of Heredity*, **98**, 468–473.
- Dupin E, Le Douarin NM (2003) Development of melanocyte precursors from the vertebrate neural crest. *Oncogene*, **22**, 3016–3023.
- Dupont A (1958) Pteridines in the Scales of Fishes. *Die Naturwissenschaften*, **45**, 267–268.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.
- Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, **491**, 756–760.
- Endler JA (1990) On the measurement and classification of colour in studies of animal colour patterns. *Biological Journal of the Linnean Society*, **41**, 315–352.
- Eriksson J, Larson G, Gunnarsson U *et al.* (2008) Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *Public Library of Science Genetics*, **4**, e1000010.

- Faircloth BC, McCormack JE, Crawford NG *et al.* (2012) Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. *Systematic Biology*, **61**, 717–726.
- Feder JL, Nosil P (2010) The Efficacy of Divergence Hitchhiking in Generating Genomic Islands during Ecological Speciation. *Evolution*, **64**, 1729–1747.
- Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in genetics*, **28**, 342–350.
- Feder JL, Gejji R, Yeaman S, Nosil P (2011) Establishment of new mutations under divergence and genome hitchhiking. *Proceedings Of The Royal Society B-Biological Sciences*, **367**, 461–474.
- Fisher S, Barry A, Abreu J *et al.* (2011) A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology*, **12**, R1.
- Forrest HS, Mitchell HK (1954) Pteridines from *Drosophila*. I. Isolation of a Yellow Pigment1. *Journal of the American Chemical Society*, **76**, 5656–5658.
- Franck MH, Schatton T (2008) Cancer stem cells and human malignant melanoma. *Pigment Cell & Melanoma Research*, **21**, 39.
- Frudakis T, Terravainen T, Thomas M (2007) Multilocus OCA2 genotypes specify human iris colors. *Human genetics*, **122**, 311–326.
- Frudakis T, Thomas M, Gaskin Z *et al.* (2003) Sequences associated with human iris pigmentation. *Genetics*, **165**, 2071–2083.
- Fu YX, Li WH (1993) Statistical Tests of Neutrality of Mutations. *Genetics*, **133**, 693–709.
- Fukamachi S, Shimada A, Shima A (2001) Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka. *Nature Genetics*, **28**, 381–385.
- Gavrilets S (2002) Sympatric speciation by sexual conflict. *Proceedings of the National Academy of Sciences*, **99**, 10533–10538.
- Giordano E, Peluso I, Rendina R, Digilio A, Furia M (2003) The clot gene of *Drosophila melanogaster* encodes a conserved member of the thioredoxin-like protein superfamily. *Molecular genetics and genomics*, **268**, 692–697.
- Goldman J (1969) Physiological Color Changes in Reptiles. *American Zoologist*, 489–504.

- Graminski GF, Jayawickreme CK, Potenza MN, Lerner MR (1993) Pigment dispersion in frog melanophores can be induced by a phorbol ester or stimulation of a recombinant receptor that activates phospholipase C. *The Journal of biological chemistry*, **268**, 5957–5964.
- Graves JAM (2006) Sex chromosome specialization and degeneration in mammals. *Cell*, **124**, 901–914.
- Graw J, Pretsch W, Löster J (2003) Mutation in Intron 6 of the Hamster Mitf Gene Leads to Skipping of the Subsequent Exon and Creates a Novel Animal Model for the Human Waardenburg Syndrome Type II. *Genetics*, **164**, 1035–1041.
- Greenwood AK, Cech JN, Peichel CL (2012) Molecular and developmental contributions to divergent pigment patterns in marine and freshwater sticklebacks. *Evolution & development*, **14**, 351–362.
- Grether GF, Hudon J, Endler JA (2001) Carotenoid scarcity, synthetic pteridine pigments and the evolution of sexual coloration in guppies (*Poecilia reticulata*). *Proceedings Of The Royal Society B-Biological Sciences*, **268**, 1245–1253.
- Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Research*, **16**, 730–737.
- Hearing VJ, Tsukamoto K (1991) Enzymatic Control of Pigmentation in Mammals. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **5**, 2902–2909.
- Hedges S, Poling L (1999) A molecular phylogeny of reptiles. *Science*, **283**, 998–1001.
- Hedges SB, Moberg KD, Maxson LR (1990) Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Molecular Biology And Evolution*, **7**, 607–633.
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.
- Hill GE (1990) Female house finches prefer colourful males: sexual selection for a condition-dependent trait. *Animal Behaviour*, **40**, 563–572.
- Hill GE, Inouye CY, Montgomerie R (2002) Dietary carotenoids predict plumage coloration in wild house finches. *Proceedings Of The Royal Society B-Biological Sciences*, **269**, 1119–1124.
- Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical population biology*, **33**, 54–78.

- Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, **97**, 222–234.
- Hoekstra HE, Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, **12**, 1185–1194.
- Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.
- Horowitz SB (1957) The effect of sulfhydryl inhibitors and thiol compounds on pigment aggregation and dispersion in the melanophores of *Anolis carolinensis*. *Experimental Cell Research*, **13**, 400–402.
- Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ (2010) Vertebrate pigmentation: from underlying genes to adaptive function. *Trends in genetics*, **26**, 231–239.
- Hugall AF, Foster R, Lee MSY (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Systematic Biology*, **56**, 543–563.
- Irizarry RA, Hobbs B, Collin F *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics (Oxford, England)*, **4**, 249–264.
- Ishida Y, David VA, Eizirik E *et al.* (2006) A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. *Genomics*, **88**, 698–705.
- Iwabe N, Hara Y, Kumazawa Y *et al.* (2005) Sister Group Relationship of Turtles to the Bird-Crocodylian Clade Revealed by Nuclear DNA-Coded Proteins. *Molecular Biology And Evolution*, **22**, 810–813.
- Janke A, Erpenbeck D, Nilsson M, Arnason U (2001) The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *Proceedings Of The Royal Society B-Biological Sciences*, **268**, 623–631.
- Jensen TA (1977) Evolution of anoline lizard display behavior. *American Zoologist*, **17**, 203–215.
- Johansson H, Surget Groba Y, Thorpe RS (2008) Microsatellite data show evidence for male-biased dispersal in the Caribbean lizard *Anolis roquet*. *Molecular Ecology*, **17**, 4425–4432.

- Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.
- Jordan RC (2008) Color-based association among heterospecifics in lake Malawi rock-dwelling cichlids. *Ethology*, **114**, 272–278.
- Jost L (2008) Gst and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.
- Katsu Y, Braun EL, Guillette LJ Jr, Iguchi T (2009) From Reptilian Phylogenomics to Reptilian Genomes: Analyses of *c-Jun* and *DJ-1* Proto-Oncogenes. *Cytogenetic and genome research*, **127**, 79–93.
- Kiefer C (2002) A class B scavenger receptor mediates the cellular uptake of carotenoids in *Drosophila*. *Proceedings of the National Academy of Sciences*, **99**, 10581–10586.
- Kiefer C, Hessel S, Lampert JM *et al.* (2001) Identification and characterization of a mammalian enzyme catalyzing the asymmetric oxidative cleavage of provitamin A. *The Journal of biological chemistry*, **276**, 14110–14116.
- Kim D, Pertea G, Trapnell C *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, **14**, R36.
- Kodric-Brown A (1985) Female preference and sexual selection for male coloration in the guppy (*Poecilia reticulata*). *Behavioral Ecology and Sociobiology*, **17**, 199–205.
- Kronforst MR, Barsh GS, Kopp A *et al.* (2012) Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell & Melanoma Research*, **25**, 411–433.
- Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Molecular Biology And Evolution*, **16**, 784–792.
- Lagos-Quintana M, Rauhut R, Yalcin A *et al.* (2002) Identification of Tissue-Specific MicroRNAs from Mouse. *Current Biology*, **12**, 735–739.
- Lamason RL, Mohideen M-APK, Mest JR *et al.* (2005) SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, **310**, 1782–1786.
- Lambert SM, Geneva AJ, Luke Mahler D, Glor RE (2013) Using genomic data to revisit an early example of reproductive character displacement in Haitian *Anolis* lizards. *Molecular Ecology*.

- Landeon EA, Badyaev AV (2012) Developmental integration of feather growth and pigmentation and its implications for the evolution of diet-derived coloration. *Journal of experimental zoology Part B, Molecular and developmental evolution*, **318**, 59–70.
- Landgraf P, Rusu M, Sheridan R *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
- Lawniczak MKN, Emrich SJ, Holloway AK *et al.* (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, **330**, 512–514.
- Lazell JD (1963) The anoles (Sauria, Iguanidae) of the Guadeloupeen Archipelago. *Bulletin of The Museum of Comparative Zoology*, **131**, 359–401.
- Lazell JD (1972) The anoles (Sauria, Iguanidae) of the Lesser Antilles. *Bulletin of the American Museum of Comparative Zoology*, **143**, 1–116.
- Le Guyader S, Maier J, Jesuthasan S (2005) Esrom, an ortholog of PAM (protein associated with c-myc), regulates pteridine synthesis in the zebrafish. *Developmental biology*, **277**, 378–386.
- Lederer E (1940) Les pigments des invertébrés. *Biological reviews of the Cambridge Philosophical Society Cambridge Philosophical Society*, **15**, 273–306.
- Lee M (1997) Reptile relationships turn turtle ... *Nature*, **389**, 245–246.
- Lee MSY, Reeder TW, Slowinski JB, Lawson R (2004) Resolving Reptile Relationships. In: *Assembling the Tree of Life* (eds Cracraft J, Donoghue MJ), pp. 451–467. Oxford University Press, Oxford.
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.
- Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Liu L, Yu L, Kubatko L, Pearl DK, Edwards SV (2009) Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics And Evolution*, **53**, 320–328.
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
- Lloyd V, Ramaswami M, Krämer H (1998) Not just pretty eyes: *Drosophila* eye-colour mutations and lysosomal delivery. *Trends in cell biology*, **8**, 257–259.

- Londos C, Gruia-Gray J, Brasaemle DL *et al.* (1996) Perilipin: possible roles in structure and metabolism of intracellular neutral lipids in adipocytes and steroidogenic cells. *International journal of obesity and related metabolic disorders*, **20 Suppl 3**, S97–101.
- Lopes SS, Yang X, Müller J *et al.* (2008) Leukocyte Tyrosine Kinase Functions in Pigment Cell Development. *Public Library of Science Genetics*, **4**, e1000026.
- Losos JB (2009) *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles*. University of California Press, Berkeley and Los Angeles, California.
- Lyson TR, Sperling EA, Heimberg AM *et al.* (2012) MicroRNAs support a turtle + lizard clade. *Biology Letters*, **8**, 104–107.
- Macedonia J, James S, Wittle L, Clark D (2000) Skin pigments and coloration in the Jamaican radiation of Anolis lizards. *Journal Of Herpetology*, **34**, 99–109.
- MacKenzie D, Nichols J, Lachman G *et al.* (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.
- Masatoshi Nei (1973) Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321.
- McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT (2011) DNA Flanking Ultraconserved Elements Provides Novel Phylogenetic Markers and Elucidates Placental Mammal Phylogeny When Combined with Species Tree Analysis. *Genome Research*, **In Review**.
- McGraw KJ, Toomey MB (2010) Carotenoid accumulation in the tissues of zebra finches: predictors of integumentary pigmentation and implications for carotenoid allocation strategies. *Physiological and biochemical zoology*, **83**, 97–109.
- McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.
- Meirmans P, Hedrick P (2011) Assessing population structure: FST and related measures. *Molecular Ecology*, **11**, 5–8.
- Menon GK (2000) Avian Epidermal Lipids: Functional Considerations and Relationship to Feathering. *Integrative and Comparative Biology*, **40**, 540–552.

- Michel AP, Sim S, Powell THQ *et al.* Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 9724–9729.
- Mills MG, Patterson LB (2009) Not just black and white: Pigment pattern development and evolution in vertebrates. *Seminars in Cell & Developmental Biology*, **20**, 72–81.
- Muñoz MM, Crawford NG, McGreevy TJ *et al.* (2013) Divergence in coloration and ecological speciation in the *Anolis marmoratus* species complex. *Molecular Ecology*, **22**, 2668–2682.
- Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268–5273.
- Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, **367**, 343–353.
- Nakayama A, Nguyen MT, Chen CC *et al.* (1998) Mutations in microphthalmia, the mouse homolog of the human deafness gene MITF, affect neuroepithelial and neural crest-derived melanocytes differently. *Mechanisms of development*, **70**, 155–166.
- Nakayama K, Fukamachi S, Kimura H *et al.* (2002) Distinctive distribution of AIM1 polymorphism among major human populations with different skin color. *Journal of human genetics*, **47**, 92–94.
- Navarro RE, Ramos-Balderas JL, Guerrero I, Pelcastre V, Maldonado E (2008) Pigment dilution mutants from fish models with connection to lysosome-related organelles and vesicular traffic genes. *Zebrafish*, **5**, 309–318.
- Neafsey DE, Lawniczak MKN, Park DJ *et al.* (2010) SNP Genotyping Defines Complex Gene-Flow Boundaries Among African Malaria Vector Mosquitoes. *Science*, **330**, 514–517.
- Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. *Annals of human genetics*, **47**, 253–259.
- NG J, Glor RE (2011) Genetic differentiation among populations of a Hispaniolan trunk anole that exhibit geographical variation in dewlap colour. *Molecular Ecology*, **20**, 4302–4317.
- Nicholson KE, Harmon LJ, Losos JB (2007) Evolution of *Anolis* lizard dewlap diversity. *Public Library of Science ONE*, **2**, e274.

- Nosil P (2008) Speciation with gene flow could be common. *Molecular Ecology*, **17**, 2103–2106.
- Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, **367**, 332–342.
- Olson V, Owens I (1998) Costly sexual signals: are carotenoids rare, risky or required? *Trends In Ecology & Evolution*, **13**, 510–514.
- Olsson M, Stuart-Fox D, Ballen C (2013) Genetics and evolution of colour patterns in reptiles. *Seminars in Cell & Developmental Biology*.
- Ord TJ, Martins EP (2006) Tracing the origins of signal diversity in anole lizards: phylogenetic approaches to inferring the evolution of complex behaviour. *Animal Behaviour*, **71**, 1411–1429.
- Ortiz E, Maldonado AA (1966) Pteridine accumulation in lizards of the genus anolis. *Caribbean Journal of Science*, **6**, 9–13.
- Ortiz E, Bächli E, Price D, Williams-Ashman HG (1963) Red pteridine pigments in the dewlaps of some anoles. *Physiological Zoology*, **36**, 97–103.
- Paemlaere EAD, Guyer C, Dobson SF (2011) A phylogenetic framework for the evolution of female polymorphism in anoles. *Biological Journal of the Linnean Society*, **104**, 303–317.
- Parichy DM, Ransom DG, Paw B, Zon LI, Johnson SL (2000) An orthologue of the kit-related gene *fms* is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, *Danio rerio*. *Development*, **127**, 3031–3044.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *Public Library of Science ONE*, **7**, e37135.
- Pointer MA, PRAGER M, Andersson S, Mundy NI (2011) A novel method for screening a vertebrate transcriptome for genes involved in carotenoid binding and metabolism. *Molecular Ecology Resources*, **12**, 149–159.
- Protas ME, Patel NH (2008) Evolution of coloration patterns. *Annual review of cell and developmental biology*, **24**, 425–446.

- Prum RO, Dufrense ER, Quinn T, Waters K (2009) Development of colour-producing β -keratin nanostructures in avian feather barbs. *Journal of the Royal Society Interface*, **6**, S253.
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
- R Core Team (2012) *R: A Language and Environment for Statistical Computing*.
- Remington DL, Thornsberry JM, Matsuoka Y *et al.* (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11479–11484.
- Rest JS, Ast JC, Austin CC *et al.* (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Molecular Phylogenetics And Evolution*, **29**, 289–297.
- Richardson J, Lundegaard PR, Reynolds NL *et al.* (2008) mc1r Pathway Regulation of Zebrafish Melanosome Dispersion. *Zebrafish*, **5**, 289–295.
- Rieppel O (1999) Turtle Origins. *Science*, **283**, 945–946.
- Rohrlich ST, Rubin R (1975) Biochemical Characterization of Crystals from the Dermal Iridophores of a Chameleon *Anolis carolinensis*. *The Journal of Cell Biology*, 635–645.
- Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
- Rosenblum EB, Hoekstra HE, Nachman MW (2004) Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution*, **58**, 1794–1808.
- Rousset D, Agnès F, Lachaume P, André C, Galibert F (1995) Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulinlike domains. *Journal of molecular evolution*, **41**, 421–429.
- Sakudoh T, Iizuka T, Narukawa J *et al.* (2010) A CD36-related Transmembrane Protein Is Coordinated with an Intracellular Lipid-binding Protein in Selective Carotenoid Transport for Cocoon Coloration. *Journal of Biological Chemistry*, **285**, 7739–7751.
- Sakudoh T, Kuwazaki S, Iizuka T *et al.* (2013) CD36 homolog divergence is responsible for the selectivity of carotenoid species migration to the silk gland of the silkworm *Bombyx mori*. *Journal of lipid research*, **54**, 482–495.

- Salzburger W, Braasch I, Meyer A (2007) Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. *BMC Biology*, **5**, 51.
- Schneider C (1996) Distinguishing between primary and secondary intergradation among morphologically differentiated populations of *Anolis marmoratus*. *Molecular Ecology*, **5**, 239–249.
- Seo T-K, Kishino H, Thorne JL (2005) Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4436–4441.
- Shedlock AM, Botka CW, Zhao S *et al.* (2007) Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 2767–2772.
- Shen X-X, Liang D, Wen J-Z, Zhang P (2011) Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles. *Molecular Biology And Evolution*, **28**, 3237–3252.
- Siepel A, Bejerano G, Pedersen JS *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, **15**, 1034–1050.
- Simpson JT, Wong K, Jackman SD *et al.* (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research*, **19**, 1117–1123.
- Soccio RE, Adams RM, Romanowski MJ *et al.* (2002) The cholesterol-regulated StarD4 gene encodes a StAR-related lipid transfer protein with two closely related homologues, StarD5 and StarD6. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6943–6948.
- Stapley J, Wordley C, Slate J (2011) No evidence of genetic differentiation between anoles with different dewlap color patterns. *Journal of Heredity*, **102**, 118–124.
- Steffen JE, McGraw KJ (2007) Contributions of pterin and carotenoid pigments to dewlap coloration in two anole species. *Comparative biochemistry and physiology Part B, Biochemistry & molecular biology*, **146**, 42–46.
- Sundvold H, Helgeland H, Baranski M, Omholt SW, Våge DI (2011) Characterisation of a novel paralog of scavenger receptor class B member I (SCARB1) in Atlantic salmon (*Salmo salar*). *BioMed Central Genetics*, **12**, 52.

- Takahashi A, Kawauchi H (2006) Evolution of melanocortin systems in fish. *General and comparative endocrinology*, **148**, 85–94.
- Tollis M, Ausubel G, Ghimire D, Boissinot S (2012) Multi-Locus Phylogeographic and Population Genetic Analysis of *Anolis carolinensis*: Historical Demography of a Genomic Model Species. *Public Library of Science ONE*, **7**, e38474.
- Trapnell C, Williams BA, Pertea G *et al.* (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, **28**, 511–515.
- Tripathi N, Hoffmann M, Willing E-M *et al.* (2009) Genetic linkage map of the guppy, *Poecilia reticulata*, and quantitative trait loci analysis of male size and colour variation. *Proceedings Of The Royal Society B-Biological Sciences*, **276**, 2195–2208.
- Tzika AC, Helaers R, Schramm G, Milinkovitch MC (2011) Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo*, **2**, 19.
- Underwood G, Williams EE (1959) The anoline lizards of Jamaica. *Bulletin of the Institute of Jamaica Science Series*, **9**, 1–48.
- Vanhooydonck B, Herrel A, Meyers JJ, Irschick DJ (2009) What determines dewlap diversity in *Anolis* lizards? An among-island comparison. *Journal Of Evolutionary Biology*, **22**, 293–305.
- Via S (2011) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions Of The Royal Society Of London Series B-Biological Sciences*, **367**, 451–460.
- Walsh N, Dale J, McGraw KJ, Pointer MA, Mundy NI (2012) Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. *Proceedings Of The Royal Society B-Biological Sciences*, **279**, 58–66.
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, **38**, e164.
- Wehrle-Haller B, Meller M, Weston JA (2001) Analysis of Melanocyte Precursors in Nf1 Mutants Reveals That MGF/KIT Signaling Promotes Directed Cell Migration Independent of Its Function in Cell Survival. *Developmental biology*, **232**, 471–483.
- Weiner L, Han R, Scicchitano BM *et al.* (2007) Dedicated epithelial recipient cells determine pigmentation patterns. *Cell*, **130**, 932–942.

- Weir BS, Hill WG (1986) Nonuniform recombination within the human beta-globin gene cluster. *American journal of human genetics*, **38**, 776–781.
- Welle M, Philipp U, Ruefenacht S *et al.* (2009) MLPH Genotype-Melanin Phenotype Correlation in Dilute Dogs. *Journal of Heredity*, **100**, S75–S79.
- Williams EE, Rand AS (1977) Species Recognition, Dewlap Function and Faunal Size. *American Zoologist*, **17**, 261–270.
- Wilson YM, Richards KL, Ford-Perriss ML, Panthier J-J, Murphy M (2004) Neural crest cell lineage segregation in the mouse neural tube. *Development*, **131**, 6153–6162.
- Wright S (1949) The genetical structure of populations. *Annals of human genetics*, **15**, 323–354.
- Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution*, **65**, 1897–1911.
- Yoder JB, Clancey E, Roches des S *et al.* (2010) Ecological opportunity and the origin of adaptive radiations. *Journal Of Evolutionary Biology*.
- Zambon AC, Gaj S, Ho I *et al.* (2012) GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics*, **28**, 2209–2210.
- Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14226–14231.
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.
- Zhong S, Joung JG, Zheng Y *et al.* (2011) High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation. *Cold Spring Harbor Protocols*, **2011**, pdb.prot5652–pdb.prot5652.
- Ziegler I (2003) The pteridine pathway in zebrafish: regulation and specification during the determination of neural crest cell-fate. *Pigment Cell Research*, **16**, 172–182.
- Ziegler I, McDonald T, Hesslinger C, Pelletier I, Boyle P (2000) Development of the pteridine pathway in the zebrafish, danio rerio. *Journal of Biological Chemistry*, **275**, 23406–18932.

Curriculum Vitae

NICHOLAS G. CRAWFORD

Email: ngcrawfo@gmail.com

Website: ngcrawford.com

Twitter: twitter.com/ngcrawford

Github: github.com/ngcrawford

EDUCATION

2007 - September 2013, Ph.D. Biology, Boston University

Advisor: Christopher Schneider

Thesis Title: Anolis Lizard Genomics and Phylogenetics

Additional Projects:

Mammal, Reptile, and Avian Phylogenomics, with Travis Glenn

Butterfly Speciation Genomics, with Sean Mullen

2004/07, M.S. Biology, San Diego State University

Advisor: Tod Reeder

Thesis: Population structure within the Plateau Striped

Whiptail (*Aspidoscelis velox* complex) a parthenogenetic lizard

Additional Project:

Phylogenetic relationships among Australian skinks of the genus

Glaphyromorphus

1997/01, B.S. Biology, Union College

Graduated Cum Laude with Academic and Departmental Honors

GRANTS AND SCHOLARSHIPS

2012 - Grand Challenges Award, Smithsonian Institute Consortia, Next Generation Phylogenetics, MJ Braun, K Wurdack, W Wcislo, J Maldonado, K Helgen, S Brady, M Cummings, TC Glenn, BC Faircloth, RT Brumeld, E Braun, JC McCormack, NG Crawford, N White. \$100,000.

2011 - Next-generation Sequencing Small Grant. Phylogenetic utility of ultra- conserved elements for the avian tree of life. MJ Braun, ND White, TC Glenn, BC Faircloth, RT Brumeld, EL Braun, JE McCormack, NG Crawford. \$10,000

2011 - Amazon Education Research Grant (aws.amazon.com), Computational Resources. BC Faircloth, NG Crawford, JE McCormack. \$10,000.

2010 - Amazon Education Research Grant (aws.amazon.com), Computational Resources. NG Crawford, BC Faircloth, TC Glenn. \$7,500.

2010 - Doctoral Dissertation Improvement Grant (NSF). NG Crawford, C Schneider. \$15,000.

2009 - Theodore Roosevelt Memorial Fund (American Museum of Natural History) labwork/fieldwork. \$1,500.

2005 - Theodore Roosevelt Memorial Fund (American Museum of Natural History) labwork/fieldwork. \$1,960.

2005 - Harry E. Hamber Memorial Scholarship, tuition. \$1,400.

2000 - IEF Grant, thesis research. \$100.

2000 - NYSEP Grant, summer stipend. \$1,700.

1999 - Booth - Ferris Grant, summer stipend. \$1,700.

PUBLICATIONS

Martha M. Munoz, Nicholas G. Crawford, Thomas J. McGreevy, Rebecca D. Tarvin, Nicholas J. Messana, Liam J. Revell, Rosanne M. Zandvliet, Juanita M. Hopwood, Elbert Mock, Andre L. Schneider, and Christopher J. Schneider. 2013. Divergence in coloration and the evolution of reproductive isolation in the *Anolis marmoratus* species complex. *Molecular Ecology*. 22(10), 2668-2682.

John E. McCormack, Michael G. Harvey, Brant C. Faircloth, Nicholas G. Crawford, Travis C. Glenn, Robb T. Brumfield. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *Public Library of Science ONE*. 8(1), e54848.

John A. St. John, Edward L. Braun, Sally R. Isberg, Lee G. Miles, Amanda Y. Chong, Jamie Gongora, Pauline Dalzell, Christopher Moran, Taisen Iguchi, Bertrand Bed'Hom, Shane C. Burgess, Amanda M. Cooksey, Todd A. Castoe, Arkhat Abzhanov, Llewellyn D.

Densmore, Miryam Venegas-Anya, Matthew J. Greenwold, Roger H. Sawyer, Federico G. Hoffmann, Nicholas G. Crawford, Jennifer C. Drew, Scott V. Edwards, Matthew K. Fujita, Jonathan M. Howard, Brant C. Faircloth, Daniel E. Janes, Shahid Yar Khan, Satomi Kohno, A.P. Jason de Koning, Stacey L. Lance, Fiona M. McCarthy, John E. McCormack, Mark E. Merchant, Daniel G. Peterson, David D. Pollock, Nader Pourmand, Brian J. Raney, Kyria A. Roessler, Jeremy R. Sanford, Carl J. Schmidt, Eric W. Triplett, Tracey D. Tuberville, Erich D. Jarvis, Louis J. Guillette Jr, Travis C. Glenn, Richard E. Green and David A. Ray. 2012. Sequencing three crocodylian genomes to illuminate the evolution of archosaurs and amniotes. *Genome Biology*. 13, 415.

Kenro Kusumi, Rob J. Kulathinal, Arhat Abzhanov, Stephane Boissinot, Nicholas G. Crawford, Brant C. Faircloth, Travis C. Glenn, Daniel E. Janes, Jonathan B. Losos, Douglas B. Menke, Steven Poe, Thomas J. Sanger, Christopher J. Schneider, Jessica Stapley, Juli Wade, Jeanne Wilson-Rawls. 2012. Developing a community-based genetic nomenclature for anole lizards. *BioMed Central Genomics*, 12(1), 554.

Nicholas G. Crawford, Brant C. Faircloth, John E. McCormack, Robb T. Brumfield, Kevin Winker, Travis C. Glenn. 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters*. 8(5), 783-786.

Brant C. Faircloth, John E. McCormack, Nicholas G. Crawford, Michael Harvey, Robb T. Brumfield, Travis C. Glenn. 2012. Ultraconserved elements anchor thousands of genetic markers for target enrichment spanning multiple evolutionary timescales. *Systematic Biology*. 61(5), 713-715.

John E. McCormack, Brant C. Faircloth, Nicholas G. Crawford, Patricia Adair Gowaty, Robb T. Brumfield, Travis C. Glenn. 2011. Ultraconserved Elements Are Novel Phylogenomic Markers that Resolve Placental Mammal Phylogeny when Combined with Species Tree Analysis. *Genome Research*. 22(4), 746-54.

Nicholas G. Crawford. 2010. SMOGD: Software for the Measurement of Genetic Diversity. *Molecular Ecology Resources*, 10: 556-557.

Nicholas G. Crawford, Jaime Zaldvar-Rae, Cris Hagen, Amanda Schable, Erica Bree Rosenblum, Jeff A. Graves, Tod W. Reeder,

Michael G. Ritchie, Travis C. Glenn. 2007. Thirteen polymorphic microsatellite DNA loci from whiptails of the genus *Aspidoscelis* (Teiidae: Squamata) and related cnemidophorine lizards.

Molecular Ecology Resources. 8: 219-223

Nicholas G. Crawford, Cris Hagen, Heather F. Sahli, Elizabeth A. Stacy, Travis C. Glenn. 2007. Fifteen polymorphic microsatellite loci from Hawaii's *Metrosideros polymorpha* (Myrtaceae: Myrtales), a model species for ecology and evolution. Molecular Ecology Resources, 8, 308-310.

Caleb R. Hickman, Maureen B. Peters, Nicholas G. Crawford, Cris Hagen, Travis C. Glenn, Christopher M. Sommers. 2008. Development and characterization of microsatellite loci in the American white pelican (*Pelecanus erythrorhynchos*). Molecular Ecology Resources, 8, 1439-1441.

Nicholas G. Crawford, Maureen B. Peters, Cris Hagen, Travis C. Glenn, Stephen K. Davis, Christopher M. Somers. 2007. Twelve polymorphic microsatellite loci from Sprague's pipit, *Anthus spragueii* (Motacillidae: Passeriformes), a threatened grassland endemic songbird. Molecular Ecology Resources, 9, 315-317.

Nicholas G. Crawford. 2007. Microsatellites in cnemidophorine lizards: their utility in investigating the landscape genetics of the plateau striped whiptail (*Aspidoscelis velox* Complex). Masters Thesis: San Diego State University.

Olga V. Tsyusko, Tracey D. Tuberville, Maureen B. Peters, Nicholas G. Crawford, Cris Hagen, Steve Weller, Ann Sakai, and Travis C. Glenn. 2007. Microsatellite markers isolated from polyploid woodsorrell (*Oxalis alpina*). Molecular Ecology Notes, 7, 1284-1286.

PROFESSIONAL EXPERIENCE

2007-2011 Teaching Fellow, Boston University: Introductory Biology, Genetics, Evolution, Animal Behavior

2007 - South Carolina, Summer, 6 months, Research Technician, supervised by Travis Glenn

2006 - South Carolina, Summer, 6 weeks, Microsatellite loci preparation at the Savannah River Ecology Laboratory, supervised

by Travis Glenn

2004/2005 - Graduate Teaching Assistant, SDSU, Introductory Biology and Introductory Zoology

2004 - Research Assistant, SDSU, Advisor: Tod Reeder

2002/03 - Research Associate, Boston Biochem

2001 - Research Associate, Pfizer Pharmaceuticals

2001 - Research Associate, Harvard Medical School

CONFERENCES, MEETINGS, AND TALKS

2013 - Invited Speaker - UMass Lowell.

Presentation Title: From Archosaurs to Anoles: Genomic Approaches to Studying Reptile Evolution

2012 - Union College Seminar Series: Invited Speaker - Union College.

Presentation Title: Genomic approaches to understanding reptile evolution

2012 - World Congress of Herpetology, Vancouver.

Presentation Title: The Genetics of Colorful Pigmentation in Anolis Lizards.

2012 - Evolutionary Genomics Super Group, Broad Institute

Presentation Title: Ultraconserved Elements as Phylogenomic Markers

2012 - MCZ Lunchtime Seminar, Harvard University

Presentation Title: Thousands of ultraconserved elements in combination with cloud computing and species-tree methods help resolve deep divergences in reptiles, mammals, and birds

2010 - The Genetics and Evolution of Animal Coloration, Radcliffe Workshop

Presentation Title: Identifying coloration genes when you can't easily do QTL mapping

2009 - Anole Symposium Harvard University

Presentation Title: *Anolis carolinensis*: Pigmentation Genetics
Poster Title: Genome scan identifies two loci associated with color polymorphism in *Anolis marmoratus*

- 2009 - Gordon Conference: Evolutionary & Ecological Functional Genomics - Tilton, New Hampshire.
Poster Title: Genetics of Colorful Pigmentation in Anolis Lizards
- 2009 - Society for Integrative and Comparative Biology Annual Meeting, Boston Massachusetts.
Presentation Title: Evolution of Dorsal Pigmentation in Anoline lizards
- 2008 - Union College Seminar Series: Invited Speaker - Union College
Presentation Title: Population Structure of the Plateau Striped Whiptail a Parthenogenetic Species of Lizard
- 2007 - Island Biogeography Symposium, Harvard University, Cambridge Massachusetts
- 2005 - American Society of Ichthyologists and Herpetologists 85th Annual Meeting, Tampa, Florida
Presentation Title: Phylogenetic relationships among Australian skinks of the genus *Glaphyromorphus*
- 2004 - American Society of Ichthyologists and Herpetologists 84th Annual Meeting, Norman, Oklahoma
- 2004 - Evolution Conference, Fort Collins, Colorado