Electronic Supplementary Material

Supplemental Tables

Table S1. Summary of diversity statistics for 5,914 SNPs and sequences (1,064 base pairs) of the mitochondrial control region for samples of bonnetheads from North Carolina (NC) and three localities along the Gulf Coast of Florida: Florida Bay (FB), Tampa Bay (TB), and Panama City (PC). *n* - sample size; π - nucleotide diversity; *H* – number of haplotypes; *h* – nucleon diversity.

SNPs	**NC**	**FB**	**TB**	**PC**
n	22	32	31	36
π	0.2996	0.3189	0.3163	0.3168
mtDNA				
n	23	24	27	25
H	7	17	17	16
h	0.719 ± 0.077	0.938 ± 0.039	0.940 ± 0.031	0.947 ± 0.029

Table S2. Distribution of haplotypes and GenBank Accession numbers for mitochondrial control region sequences from samples of bonnetheads off North Carolina (NC) and three locations along the Gulf Coast of Florida: Florida Bay (FB), Tampa Bay (TB) and Panama City (PC).

	NC	FB	TB	PC	Genbank Accession #
H1	6	1	2	2	KT031755
H2	11	1	6	3	KT031756
H3	1	-	-	-	KT031757
H4	1	-	-	-	KT031758
H5	1	-	-	-	KT031759
H6	2	6	2	5	KT031760
H7	1	-	-	-	KT031761
H8	-	-	-	1	KT031762
H9	-	-	-	2	KT031763
H10	-	-	-	1	KT031764
H11	-	-	-	1	KT031765
H12	-	-	-	2	KT031766
H13	-	-	-	1	KT031767
H14	-	-	1	1	KT031768
H15	-	-	-	1	KT031769
H16	-	-	-	1	KT031770
H17	-	2	3	1	KT031771
H18	-	1	1	1	KT031772
H19	-	-	-	1	KT031773
H20	-	-	-	1	KT031774
H21	-	-	1	-	KT031775
H22	-	-	2	-	KT031776
H23	-	-	1	-	KT031777
H24	-	1	-	-	KT031778
H25	-	-	1	-	KT031779
H26	-	1	-	-	KT031780
H27	-	1	-	-	KT031781
H28	-	1	-	-	KT031782
H29	-	1	-	-	KT031783
H30	-	-	1	-	KT031784
H31	-	-	1	-	KT031785
H32	-	1	-	-	KT031786
H33	-	1	-	-	KT031787
H34	-	1	-	-	KT031788
H35	-	-	1	-	KT031789

H36	-	-	1	-	KT031790
H37	-	1	-	-	KT031791
H38	-	1	-	-	KT031792
H39	-	2	-	-	KT031793
H40	-	-	1	-	KT031794
H41	-	-	1	-	KT031795
H42	-	1	-	-	KT031796
H43	-	-	1	-	KT031797
H44	-	1	-	-	KT031798
Total	23	25	27	25	

Table S3. Results of standard least squares regression of allele frequencies at outlier loci by latitude: loci are organized as bi-allelic and multi-allelic. Values of $r^2 > 0.90$ and $P < 0.05$ are italicized and bolded. %X and %Y are percentage contribution to abscissa and ordinate from DAPC analysis, using k-means clustering.

Haplotype	r^2	***P*-value**	% X	%Y
Bi-allelic				
E28275:1/2	***0.93***	***0.037***	7.54	0.08
E29236:1/2	***0.92***	***0.042***	0.42	0.62
E43805:1/2	***0.90***	***0.050***	6.22	0.08
E53310:1/2	0.85	0.075	2.56	1.10
E61036:1/2	0.49	0.299	0.02	0.16
E64126:1/2	***0.91***	***0.048***	0.56	0.02
E65274:1/2	0.76	0.130	0.54	1.80
E66074:1/2	***0.92***	***0.043***	8.94	0.04
E68107:1/2	0.76	0.127	2.64	3.56
E82240:1/2	***0.90***	0.053	7.50	0.70
E87901:1/2	0.83	0.089	0.46	4.36
E92875:1/2	***0.93***	***0.035***	2.68	0.00
E94551:1/2	***0.92***	***0.041***	8.28	0.02
E95266:1/2	***0.90***	***0.049***	0.90	0.14
E102269:1/2	0.84	0.084	0.02	10.58
E106435:1/2	***0.92***	***0.041***	8.28	0.02
E107376:1/2	0.61	0.222	0.18	1.28
E113131:1/2	0.86	0.073	2.10	1.24
E117105:1/2	0.86	0.071	3.52	0.38
E167424:1/2	0.87	0.069	2.04	1.76
E195370:1/2	0.83	0.089	3.54	1.12
Multi-allelic				
E52101:1/…	0.67	0.179	0.22	0.26
:2/…	0.29	0.457	0.07	13.98
:3/…	0.06	0.763	0.55	10.46
E69589:1/…	0.76	0.126	0.36	2.44

:2/…	0.87	0.070	0.32	3.18
:3/…	***0.92***	***0.042***	0.00	0.05
E73988:1/…	***0.92***	***0.038***	2.81	0.09
:2/…	***0.92***	***0.043***	3.48	0.07
:3/…	0.84	0.085	0.04	0.31
E75833:1/…	0.17	0.593	0.05	4.02
:2/…	**0.90**	0.054	1.05	1.14
:3/…	**0.94**	***0.029***	0.62	0.88
E94553:1/…	0.87	0.069	2.31	0.21
:2/…	0.79	0.112	0.58	0.07
:3/…	***0.92***	***0.042***	0.57	0.53
E110379:1/…	0.35	0.405	0.24	18.46
:2/…	0.89	0.057	0.86	7.43
:3/…	***0.93***	***0.034***	2.08	2.92
E71001:1/…	***0.92***	***0.039***	1.09	0.07
:2/…	0.73	0.146	0.05	1.18
:3/…	0.79	0.109	0.63	1.07
:4/…	0.67	0.182	0.00	0.04
E109425:1/…	***0.95***	***0.028***	2.35	1.15
:2/…	***0.92***	***0.038***	4.37	0.11
:3/…	0.45	0.331	0.34	0.48
:4/…	0.02	0.862	0.00	0.00
E131866:1/…	***0.90***	0.052	3.91	0.01
:2/…	0.89	0.055	0.04	0.06
:3/…	***0.93***	***0.037***	1.92	0.24
:4/…	0.72	0.150	0.15	0.03

Table S4. Results of analysis of molecular variance (AMOVA) for all three data sets. Data include % of variation (%), degrees of freedom (df), and sum of squares (SS).

AMOVA using Outlier Loci (O-SNPs):

Source of Variation	**Nested in**	**%**	**df**	**SS**	F_{ST}	***P*-value**
Among Individuals	Population	72.93	236	1245.48	--	--
Among Population	--	27.07	3	365.10	0.271	<0.00001

AMOVA using Neutral Loci (N-SNPs):

Source of Variation	**Nested in**	**%**	**df**	**SS**	F_{ST}	***P*-value**
Among Individuals	Population	99.21	236	163871.48	--	--
Among Population	--	0.79	3	3068.96	0.008	<0.00001

AMOVA using mtDNA haplotypes:

Source of Variation	**Nested in**	**%**	**df**	**SS**	Φ_{ST}	***P*-value**
Among Individuals	Population	92.23	95	74.52		--
Among Populations	--	7.77	3	7.26	0.078	<0.00001

Table S5. Results of BLAST search for sequence similarity of SNP containing loci.

Locus	Description	E-value	PI	Accession #	Hit start	Hit end
Di-Allelic						
E28275_L100	zebrafish DNA clone CH211-168G2 in linkage group 9	1.06E-03	83.7	BX005152	12226	12274
	zebrafish DNA clone DKEY-159B16 in linkage group 9	1.06E-03	83.7	CR407600	70212	70164
	Mustelus manazo DNA, HE1 SINE clone: Mm 3	3.72E-03	70.8	AB027718	181	86
E29236_L100	*Triakis scyllium* TNFRSF6B gene for tumor necrosis factor	9.96E-17	83.5	AB282596	1771	1681
	Carcharhinus leucas microsatellite C105	3.48E-16	83.0	KJ916108	55	147
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	1.21E-15	82.4	AB174766	2935	3025
E43805_L100	*Ginglymostoma cirratum* clone GC_BA-557	1.69E-07	84.7	AC164927	65082	65024
	Ginglymostoma cirratum clone GC_BA-557B6	1.69E-07	84.7	AC164927	83026	82978
	Carcharhinus plumbeus RAG1 (partial) and RAG2 (complete)	1.69E-07	84.7	AY172838	8908	8966
E53310_L100	*Carcharhinus isodon* clone *Cis*172 microsatellite	1.06E-03	89.7	JQ365996	56	94
	Predicted: *Equus caballus* Zinc finger protein 226 (ZNF226)	1.3E-02	87.5	XM_005596327	2151	2112
	Predicted: *Equus caballus* Zinc finger protein 226 (ZNF226)	1.3E-02	87.5	XM_005596328	1916	1877
E61036_L100	None					
E64126_L100	None					
E65274_L142	*Mustelus manazo* DNA, HE1 SINE, clone:Mm 2	6.29E-13	85.9	AB027717	38	108
	Triakis scyllium IL-1 gene for interleukin-1β	7.66E-12	86.6	AB074142	3145	3211
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	2.67E-11	84.5	AB174766	3231	3161
E66074_L100	None					
E68107_L100	Predicted*: Latimeria chalumnae* Zinc finger protein 850-like (LOC1 02360397)	2.67E-11	80.7	XM_005995770	1396	1314
	Predicted: *Colius striatus* Zinc finger protein 501-like (LOC10455 9532)	1.14E-09	80.0	XM_010204559	482	403
	Predicted: *Xenopus* (*silurana) tropicalis* Zinc finger protein 84-like (LOC101730679)	3.97E-09	76.6	XM_004919267	571	478
E87901_L109	*Carcharhinus sorrah* microsatellite *Cs*08	5.88E-26	87.4	AY545211	752	643
	Carcharhinus plumbeus T cell receptor gamma (TCRG)	2.05E-25	86.4	FJ854492	17993	18102
E92875_L100	*Stegastes partitus* Zinc finger and scan domain-protein 2-like (LOC103371773)	3.72E-03	86.0	XM_008301224	1162	1120
	Stegastes partitus Zinc finger and scan domain-protein 2-like (LOC103371773)	3.72E-03	86.0	XM_008301232	782	740
E94551_L100	*Mustelus manazo* DNA, HE1 SINE clone:Mm2	7.66E-12	92.6	AB027717	65	12
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	7.66E-12	92.6	AB174766	3204	3257

	Mustelus manazo DNA, HE1 SINE clone:Mm 3	3.26E-10	90.7	AB027718	65	12
E95266_L148	*Callorhinchus milii* clone P02H01.kidney.K065	1.3E-02	91.4	JX052830	4	38
	Callorhinchus milii clone, NADH dehydrogenase (Ubiquinone) 1 beta subcomplex	1.3E-02	91.4	JX209076	25	59
	Callorhinchus milii NADH Dehydrogenase (Ubiquinone) 1 beta subcomplex	1.3E-02	91.4	NM_001292677	1	35
E102269_L100	None					
E106435_L100	*Mustelus manazo* DNA, HE1 SINE, clone:Mm 2	3.71E-22	86.9	AB027717	239	144
	Mustelus manazo DNA, HE1 SINE, clone:Mm 4	3.71E-22	88.9	AB027719	231	143
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	3.71E-22	87.9	AB174766	3030	3125
E107376_L100	*Carcharhinus plumbeus* T cell receptor gamma (TCRG)	1.06E-03	72.3	FJ854492	25169	25069
E113131_L101	*Ginglymostoma cirratum* Zinc finger protein 112, β2m , βrd2 genes	3.72E-03	84.8	AB571627	68553	68598
E117105_L100	None					
E195370_L100	*Brugia pahangi* genome assembly B	3.72E-03	93.9	LK964241	31402	31434
Multi-allelic						
E52101_L100	*Scyliorhinus canicula* Cluster_HOXD sequence	5.00E-15	83.0	FQ032660	58525	58612
	Mustelus manazo DNA, HE1 SINE, clone:Mm 2	2.00E-12	81.0	AB027717	52	136
	Carcharhinus plumbeus RAG1 (partial) and RAG2 (complete)	1.0E-10	80.0	AY172838	3341	3425
E69589_L100	*Carcharhinus plumbeus* T cell receptor gamma (TCRG)	4.0E-09	83.0	FJ854492	18165	18105
	Carcharhinus sorrah microsatellite CS08 sequence	6.00E-07	90.0	AY545211	601	648
	Carcharhinus plumbeus Ig lambda light chain gene, COMPLETE CDS	6.00E-07	90.0	U34992	6243	6290
E75833_L160	*Sphyrna lewin* GRLN Gene for preproghrelin, COMPLETE CDS	3.26E-10	79.5	AB254130	3838	3925
	Mustelus manazo DNA, HE1 SINE, clone:Mm 2	3.97E-09	78.7	AB027717	57	145
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	4.83E-08	77.5	AB174766	3212	3124
E82240_L100	*Homo sapiens* BAC clone RP11-334C6 from chromosome 7	1.3E-02	84.4	AC073418	55859	55815
	Pan troglodytes BAC clone CH251-734F1 from chromosome 7	1.3E-02	84.4	AC190230	179380	179336
	Pan troglodytes BAC clone CH251-541E24 from chromosome 7	1.3E-02	84.4	AC192728	31156	31112
E94553_L100	*Botryotinia fuckeliana* T4 supercontig_34_1 genomic supercontig	1.3E-02	84.0	FQ790278	92859	92814
	Botryotinia fuckeliana B05.10 hypothetical protein (BC1G_13291)	1.3E-02	84.0	XM_001548305	1089	1044
	Torpedo marmorata mRNA fragment for acetylcholinesterase c-term	4.53E-02	96.6	X13173	1745	1773
E71001_L100	None					
E73988_L143	*Scyliorhinus canicula* cluster_HOXBS	2.00E-14	79.0	FQ032659	65153	65252
	Triakis scyllium IL-1 gene for interleukin-1β	2.00E-13	79.0	AB074142	899	1002
	Scyliorhinus canicula cluster_HOXD sequence	1.00E-10	78.0	FQ032660	15011	15107

E109425_L100	None					
E110379_L124	None					
E131866_L100	*Triakis scyllium* IL-1 gene for interleukin-1β	2.34E-18	81.7	AB074142	1087	984
	Triakis scyllium MIP3 gene for macrophage inflammatory protein-3α	9.96E-17	81.6	AB174766	3443	3540
	Ginglymostoma cirratum clone GC_Ba-678C3	4.24E-15	80.6	AC165195	14263	14360
E167414_L100	*Callorhinchus milii* gamma-aminobutyric acid type B receptor	3.05E-04	83.0	XM_007884901	1038	1090

Supplemental Experimental Procedures

ddRAD Library Methods

DNA was extracted using Mag-Bind Tissue DNA kits (Omega Bio-Tek) and digested with *Eco*RI and *Msp*I. A barcoded adapter was ligated to *Eco*RI restriction sites and a common adapter was ligated to *Msp*I restriction sites, using equimolar quantities of each digested sample. Samples were then pooled into four 'index' libraries consisting of ~34 individuals each and size selected using a Pippin Prep DNA size selection system (Sage Science Inc.). Fragments were selected using a mean size of 375 bp, with a 'tight' selection window (± 37 bp). Illumina flow-cell adapter sequences and index-specific identifiers were added to each index library, using 12 cycles of PCR.

mtDNA Sequencing

Thirty-microlitre PCR reactions contained 1X reaction buffer (pH 8.5), 2 mM MgCl2, 0.25 mM of each dNTP, 250 pmol of each primer, 0.05 U/μL *Taq* polymerase and 1.0 μL of template. Reaction conditions consisted of initial denaturation at 95°C for 2 min followed by 35 cycles at 94°C for 30 s, 54° for 60 s, and 72°C for 65 s, and a final extension at 72°C for 10 min. Amplified products were sent to Beckman Coulter to be cleaned and sequenced bi-directionally.

Bioinformatic Analysis

The combination of paired-end (PE) and single-end (SE) libraries called for some customization to the default *dDocent* pipeline [S1] (Puritz et al. 2014). PE read files were placed in a working directory and a modified version of *dDocent* (version 1.0) was run with a cutoff value of 2 and a clustering % of 0.95. There were three modifications for reference contig assembly to help deal with a large repetitive genome.

1. During the clustering command of *Rainbow* [S2] (Chong et al. 2012), the mismatch parameter was changed to 2 and the –L command was implemented.
2. During the div command of *Rainbow* [S2] (Chong et al. 2012), the minimum frequency for a new variant was changed from 0.2 to 0.05
3. During the merge command of *Rainbow* [S2] (Chong et al. 2012), the minimum number of reads to assemble was lowered from 5 to 3 and the maximum number of divided clusters to merge and the maximum number of reads to assemble was raised from 300 to 500.

After reference assembly was completed, SE reads were returned to the working directory. The read mapping portion of *dDocent* (version 1.0) was modified again to help deal with a large repetitive genome. The default clipping penalties were changed to (20,5), the mismatch parameter was lowered from 4 to 3, and the gap opening penalty was lowered from 6 to 5. Additionally, reads with more than 20% clipping were removed with AWK after mapping. These settings would enhance the ability for highly polymorphic reads to map to the reference, but also remove reads that have only have a small matching portion. Afterwards, default values of *dDocent* (version 1.0) were used to call to call variants. Raw variant calls were subjected to several filtering steps to reduce false positives. A script to reproduce the filtering steps can be found at (Dryad XXXX). Raw variants were filtered sequentially via VCFtools [S3](Danecek et al. 2011) or custom bash scripts, using the following steps:

1. Loci were removed that had a minor allele count of less than 2, a PHRED quality score of less than 20, and a call rate of less than 50%.
2. All genotypes with less than 5 reads were changed to missing.
3. Loci were removed that now had less than a 75% call rate, and a minor allele frequency of less than 0.025.
4. 13 individuals (out of 134) were removed for having more than 26.5% missing data, using the script filter_missing_ind.sh (https://github.com/jpuritz/dDocent/blob/master/scripts/filter_missing_ind.sh).
5. The highly related individual (as described in main text; LK_00700) was removed. Loci called in less than 97.5% of individuals and with a minor allele frequency of less than 5% were removed.

After this point, variant calls were filtered using a custom script (FB_filters_Bhead.sh; Dryad: XXXX) that utilizes vcflib (https://github.com/ekg/vcflib) and VCFtools [S3](Danecek et al., 2011) to filter loci based on FreeBayes INFO criteria and depth:

1. Loci were removed if the average allele balance at heterozygous genotypes was less than 28% (i.e., if a genotype had 100X coverage, there would have to be an average of 28 or more reads from the alternate allele across all heterozygous genotypes). Additionally, if the quality sum of the reference or alternate allele was 0, the locus was removed. This removes sites that have a large portion of spurious heterozygous genotype calls.

2. Loci were then removed if the quality score was less than half of the total depth. With FreeBayes, excessive depth can give inflated quality scores.
3. Loci were removed if the ratio between the mean mapping quality of the alternate and reference allele was less than 0.9 or more than 1.05.
4. Loci were then removed if the majority of reads did not come from only one read orientation. Our insert size was much larger than our PE read lengths, so true RAD loci should not have forward and reverse reads that overlap.
5. Loci also were removed based on the status of properly paired reads. True variants should have reads coming from all properly paired reads, or only from reads that are not properly paired (some RAD loci do not assemble well for the PE read, leaving only forward reads). However, false variants tend to have properly paired reference reads and not properly paired alternate reads. Loci were retained if more than 0.05% of reference reads were properly paired and less than 0.05% of alternate reads were properly paired and vice versa.
6. Of the remaining loci, the average depth (and standard deviation) across all individuals was calculated. Loci that have a depth greater than the average depth plus on SD are removed if the quality score is less than 2 times the depth. This filter is based off results reported in (arXiv:1404.0929v1) by Li [S4] (2014).
7. Only loci that were in the bottom 90% of mean depth (less than 79.94) were kept to remove any possible paralogs or repetitive genomic regions.

After these two filter steps, loci were filtered based on locality-based tests of Hardy Weinberg equilibrium. Loci were removed that had a *P* value of less than 0.001 in at least 25% of the populations, using the script filter_hwe_by_pop.pl (https://github.com/jpuritz/dDocent/blob/master/scripts/filter_hwe_by_pop.pl). Variant calls were then decomposed into SNP and INDEL calls, using vcflib; INDELS were then removed with VCFtools to produce a VCF file of SNP only calls. SNPs were then filtered again for paralogs by removing loci that had more than 3 SNPs within 5 bp and loci that had had more than 4 SNPs within 17 bp. This final set of SNP calls was used for all subsequent analyses.

Supplemental References

1. Puritz, J.B., Hollenbeck, C.M., and Gold, J.R. (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2, e431.
2. Chong, Z., Ruan, J., and Wu, C.-I. (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. Bioinformatics 28, 2732-2737
3. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011) The variant call format and VCFtools. Bioinformatics 27, 2156-2158.
4. Li, H. (2014) Towards better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics 30, 2843-2851.