

Genetic structure across isolated Virginia populations of the endangered candy darter (*Etheostoma osburni*)

Kathryn E. McBaine^{1,2}, Paul L. Angermeier^{3,1}, Eric M. Hallerman^{1*}

¹ Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0321, USA; katimc9@vt.edu, ehallerm@vt.edu

² Current Affiliation: Idaho Department of Fish and Game, 1414 E. Locust Lane, Nampa, ID 83686, USA

³ U.S. Geological Survey, Virginia Cooperative Fish and Wildlife Research Unit; Department of Fish and Wildlife Conservation, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0321, biota@vt.edu

* Correspondence: ehallerm@vt.edu; Tel.: +001-540-553-5873

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Table S1. Site locations for Cripple Creek, Dismal Creek, Stony Creek, and Laurel Creek for all surveys. Coordinates are in decimal degrees.

Stream	Location within stream	Years sampled	Latitude	Longitude
Cripple	Lower	2016 & 2017	36.9	-80.9
	Center	2016 & 2017	36.9	-81.0
	Upper	2016 & 2017	36.9	-81.0
Dismal Creek	Below Dismal Falls	2017	37.2	-80.9
	Above Dismal Falls 1	2017	37.2	-80.9
	Above Dismal Falls 2	2017	37.2	-80.9
Stony Creek	Lower 1	2016, 2017, 2018	37.4	-80.7
	Lower 2	2017 & 2018	37.4	-80.7
	Center 1	2017 & 2018	37.4	-80.6
	Center 2	2016, 2017, 2018	37.4	-80.6
	Upper	2017 & 2018	37.4	-80.6
Laurel Creek	Lower 1	2017 & 2018	37.2	-81.1
	Lower 2	2017 & 2018	37.3	-81.1
	Center	2016, 2017, 2018	37.3	-81.1
	Upper 1	2017 & 2018	37.3	-81.1
	Upper 2	2017 & 2018	37.3	-81.1

Table S2. Locus-by-locus genetic diversity metrics for four Virginia populations of candy darter, *Etheostoma osburni*, across 12 microsatellite loci. N = number of fish sampled, A = mean number of alleles per locus, A_r = allelic richness, H_o = mean observed heterozygosity, H_e = mean expected heterozygosity, Allelic Range = mean difference between sizes of largest and smallest alleles at a particular locus, M = ratio of number of alleles observed at a locus to number of alleles possible between the largest and smallest alleles (Garza and Williamson 2001).

Stream	Locus	N	A	A_r	H_o	H_e	Allelic Range	M
Cripple Creek	<i>D116</i>	98	11	7.61	0.82	0.81	284.00	0.04
	<i>C124</i>	98	9	8.09	0.69	0.84	80.00	0.11
	<i>C6</i>	98	3	3.00	0.59	0.64	8.00	0.33
	<i>D107</i>	98	11	7.51	0.67	0.76	268.00	0.04
	<i>C208</i>	98	3	2.97	0.43	0.45	16.00	0.18
	<i>C112</i>	98	6	5.23	0.76	0.77	36.00	0.16
	<i>C207</i>	98	5	4.31	0.63	0.63	16.00	0.29
	<i>C117</i>	98	6	4.69	0.61	0.65	264.00	0.02
	<i>C3</i>	98	3	2.00	0.29	0.45	212.00	0.01
	<i>D10</i>	98	9	6.99	0.78	0.83	282.00	0.03
	<i>D108</i>	98	7	5.69	0.78	0.80	232.00	0.03
	<i>D11</i>	98	7	5.21	0.63	0.77	234.00	0.03
Dismal Creek	<i>D116</i>	50	5	4.00	0.52	0.67	292.00	0.02
	<i>C124</i>	50	10	8.66	0.60	0.82	348.00	0.03
	<i>C6</i>	50	2	1.00	0.00	0.22	248.00	0.01
	<i>D107</i>	50	5	3.90	0.44	0.73	272.00	0.02
	<i>C208</i>	50	2	1.00	0.00	0.08	234.00	0.01
	<i>C112</i>	50	2	1.00	0.00	0.08	178.00	0.01
	<i>C207</i>	50	3	2.00	0.20	0.25	216.00	0.01
	<i>C117</i>	50	2	1.00	0.00	0.08	260.00	0.01
	<i>C3</i>	50	4	2.85	0.36	0.67	216.00	0.02
	<i>D10</i>	50	4	3.00	0.24	0.66	262.00	0.02
	<i>D108</i>	50	5	3.77	0.48	0.71	204.00	0.02
	<i>D11</i>	50	3	2.00	0.20	0.56	238.00	0.01
Laurel Creek	<i>D116</i>	94	14	9.55	0.79	0.86	332.00	0.04
	<i>C124</i>	94	10	7.33	0.49	0.79	320.00	0.03
	<i>C6</i>	94	5	3.95	0.62	0.69	260.00	0.02
	<i>D107</i>	94	11	7.89	0.60	0.79	300.00	0.04
	<i>C208</i>	94	2	1.60	0.04	0.04	12.00	0.15
	<i>C112</i>	94	6	5.09	0.70	0.66	28.00	0.21

	<i>C207</i>	94	3	2.97	0.43	0.36	16.00	0.18
	<i>C117</i>	94	4	3.99	0.60	0.61	28.00	0.14
	<i>C3</i>	94	3	2.00	0.38	0.40	212.00	0.01
	<i>D10</i>	94	7	5.61	0.57	0.77	274.00	0.03
	<i>D108</i>	94	9	7.29	0.68	0.73	236.00	0.04
	<i>D11</i>	94	10	7.77	0.79	0.84	246.00	0.04
Stony Creek	<i>D116</i>	94	14	6.35	0.79	0.86	332.00	0.04
	<i>C124</i>	94	10	9.78	0.49	0.79	320.00	0.03
	<i>C6</i>	94	5	2.00	0.62	0.69	260.00	0.02
	<i>D107</i>	94	11	8.80	0.60	0.79	300.00	0.04
	<i>C208</i>	94	2	2.99	0.04	0.04	12.00	0.15
	<i>C112</i>	94	6	4.35	0.70	0.66	28.00	0.21
	<i>C207</i>	94	3	3.34	0.43	0.36	16.00	0.18
	<i>C117</i>	94	4	2.00	0.60	0.61	28.00	0.14
	<i>C3</i>	94	3	2.00	0.38	0.40	212.00	0.01
	<i>D10</i>	94	7	4.75	0.57	0.77	274.00	0.03
	<i>D108</i>	94	9	3.00	0.68	0.73	236.00	0.04
	<i>D11</i>	94	10	5.38	0.79	0.84	246.00	0.04

Table S3. Mean log probabilities, $\text{Ln}P(D|K)$, for each given number of multilocus genotypic clusters (K) provided by STRUCTURE Bayesian cluster analysis. Reps = replicate runs and Stdev = standard deviation. N/A = no metric relevant for this value of K .

K	Reps	Mean $\text{Ln}P(K)$	Stdev $\text{Ln}P(K)$	$\text{Ln}'(K)$	$ \text{Ln}''(K) $	Delta K
1	10	-6904	0.434102	N/A	N/A	N/A
2	10	-6027.1	90.26245	876.9	220.68	2.44487
3	10	-5370.9	39.2259	656.22	124.39	3.171119
4	10	-4839	0.254078	531.83	511.17	2011.8637
5	10	-4818.4	11.69673	20.66	132.55	11.332229
6	10	-4930.3	44.60997	-111.89	25.08	0.562206
7	10	-5017.1	22.72366	-86.81	85.32	3.754677
8	10	-5018.6	47.16937	-1.49	N/A	N/A

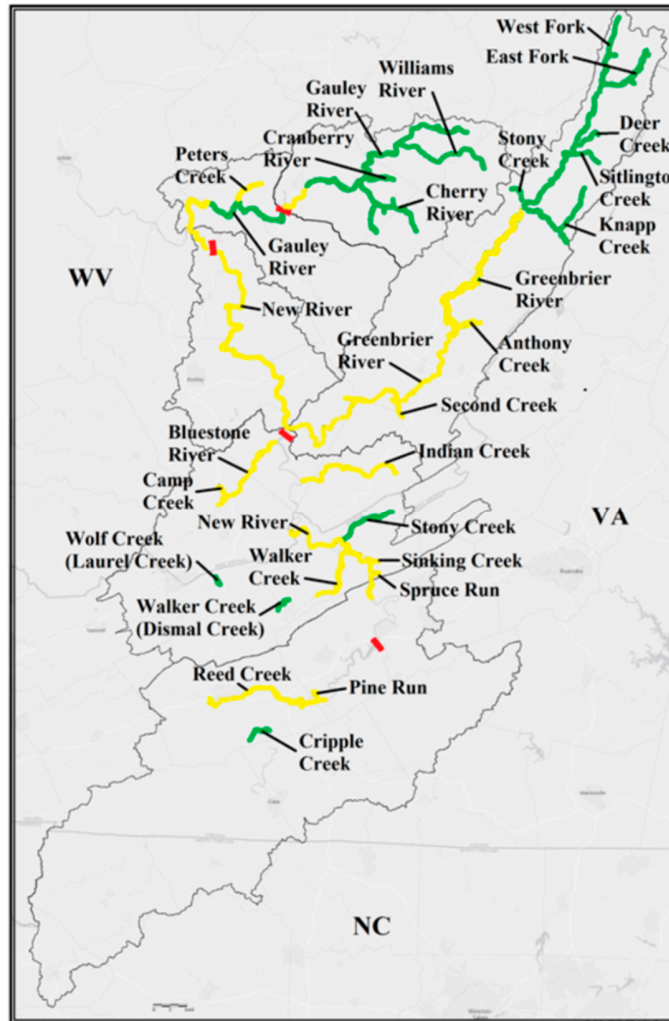
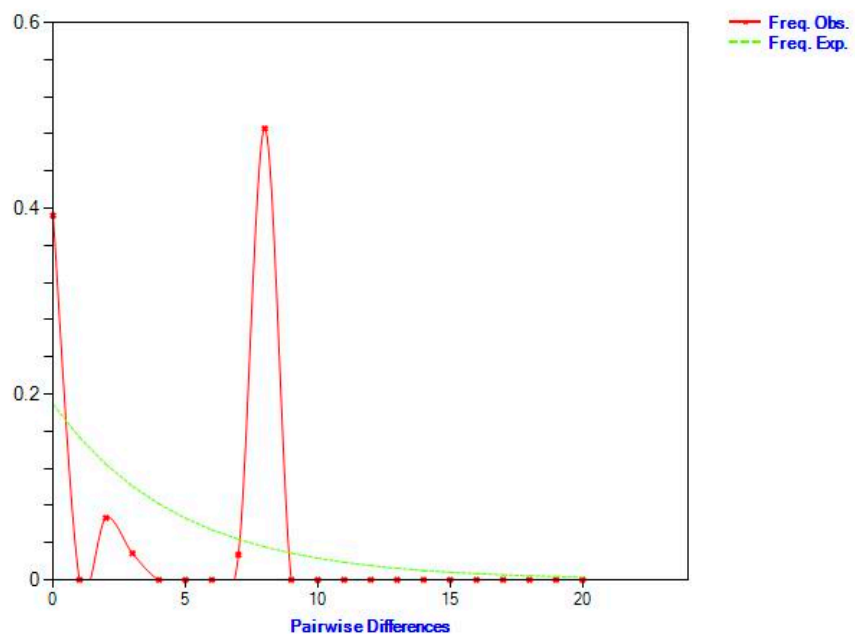
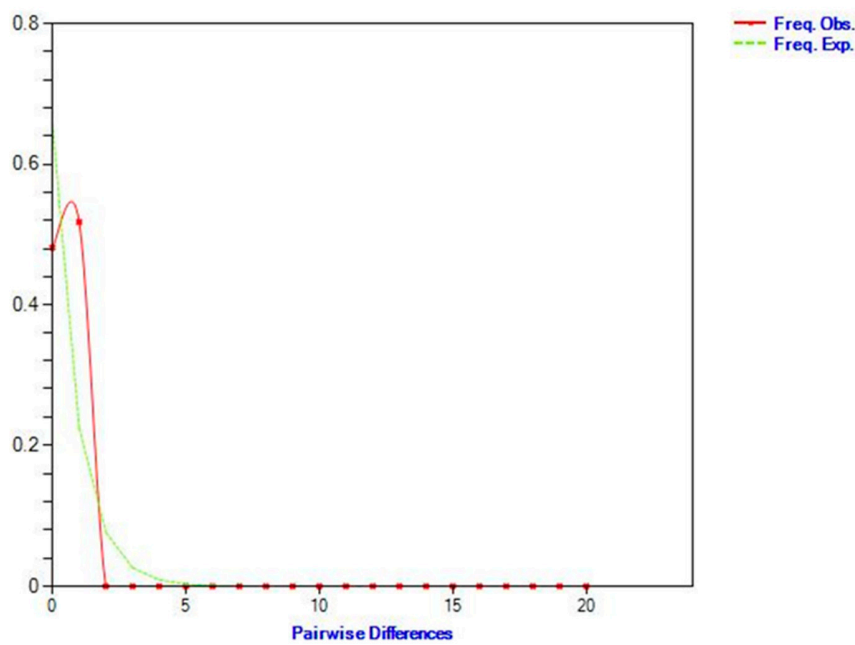


Figure S1. Current and historical distribution of the candy darter, *Etheostoma osburni* (USFWS 2018). Green indicates extant candy darter populations; yellow indicates historical or extirpated populations. Red lines are major dams that present barriers to fish movement.

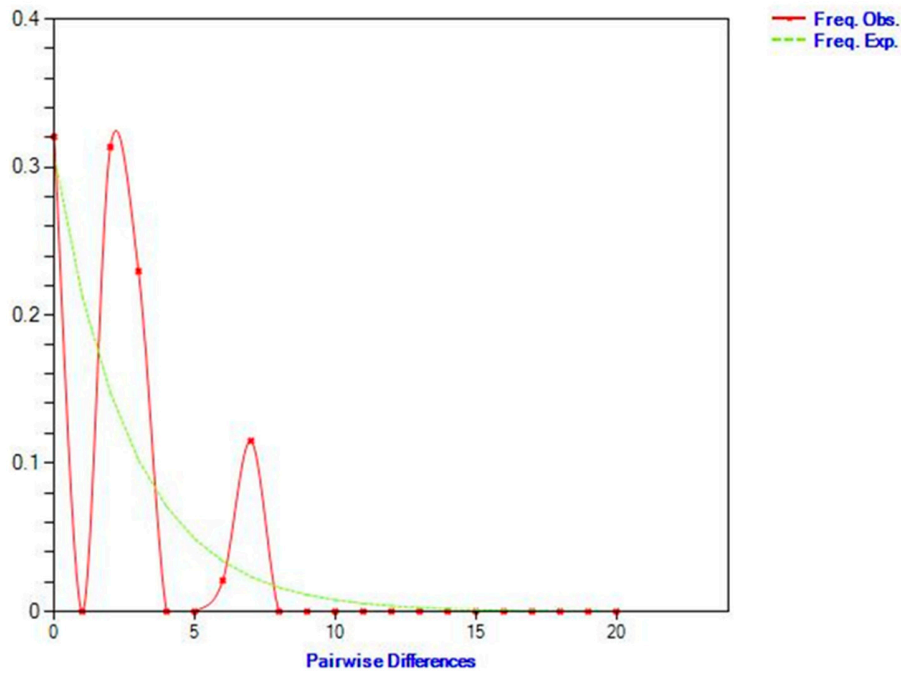
a.



b.



c.



d.

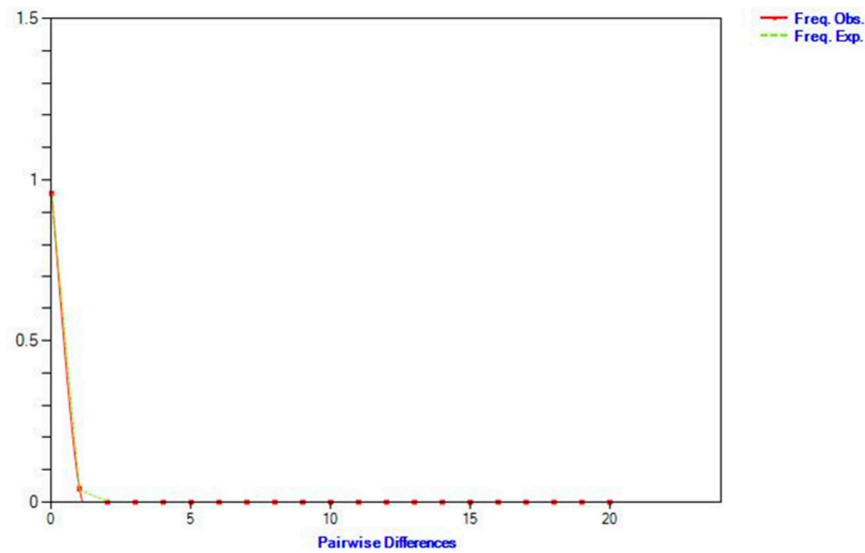


Figure S3. DNA sequence mismatch distributions of pairwise nucleotide differences in the a) Cripple Creek, b) Dismal Creek, c) Laurel Creek, and d) Stony Creek populations of candy darter, *Etheostoma osburni*. Freq. Obs. = Frequency observed; Freq. Exp. = Frequency expected.

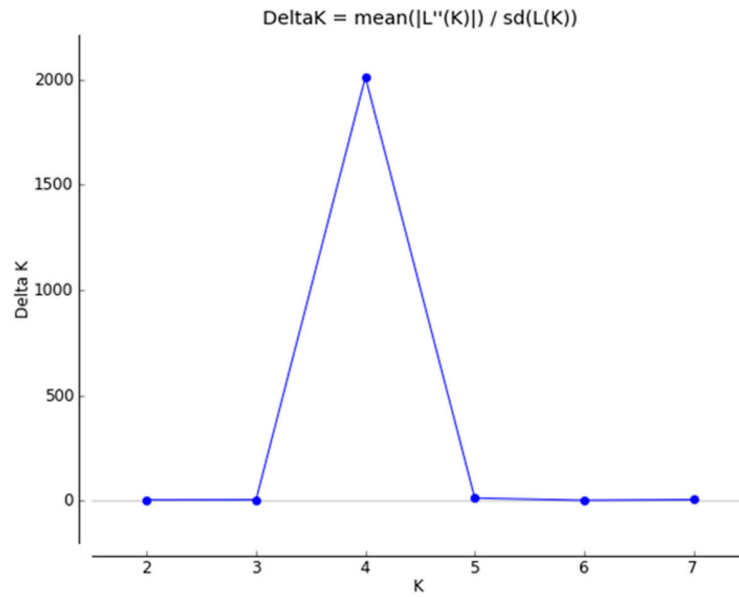


Figure S4. Results from application of the Evanno et al. (2005) method for determining the best-supported number of genetic clusters (K) using 12 loci in STRUCTURE. Clusters represent similar multi-locus genotypes of individual Candy Darters, *Etheostoma osburni*, regardless of point of collection. DeltaK = change in log probabilities of each K , $\text{mean}(|L''(K)|)$ = mean log likelihood among replicates for each value of K , $\text{sd}(L(K))$ = standard deviation of log likelihoods among replicates for the same K .