

Supporting information for

Against the odds: recent hybrid zones between mangrove killifish species with different mating systems

Waldir M. Berbel-Filho^{1,2*}, Andrey Tatarenkov³, George Pacheco⁴, Helder M. V. Espírito-Santo⁵, Mateus G. Lira⁶, Carlos Garcia de Leaniz², John C. Avise³, Sergio M. Q. Lima⁶, Carlos M. Rodríguez-López⁷, and Sofia Consuegra^{2*}

¹ Department of Biology, University of Oklahoma, Norman, OK, USA. (present address)

² Department of Biosciences, College of Science, Swansea University, Swansea, UK.

³ Department of Ecology and Evolutionary Biology, University of California, Irvine, USA.

⁴ Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, 1353 Copenhagen, Denmark.

⁵ Núcleo de Ecologia Aquática e Pesca da Amazônia, Universidade Federal do Pará, Belém, Brazil.

⁶ Laboratório de Ictiologia Sistemática e Evolutiva, Departamento de Botânica e Zoologia, Universidade Federal do Rio Grande, Natal, Brazil.

⁷ Environmental Epigenetics and Genetics Group, Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, Lexington, KY, USA.

*Corresponding authors: waldirmbf@gmail.com; s.consuegra@swansea.ac.uk

Library preparation

200ng of genomic DNA were digested using a *EcoRI* (cutsite: GAATTC) and *HpaII* (cutsite: CCGG). Digested DNA was ligated to individually barcoded adapters with a *HpaII* cut site overhang and a common *EcoRI* Y adapter. Ligation products were individually cleaned to remove excess of adapters using Agencourt AMPure XP purification system (#A63880, Beckman Coulter, Brea, CA, USA) at a v/v ratio of 0.85 following the manufacturer's instructions. A single library was produced by pooling 20ng of DNA from each sample and amplified in eight separate PCR reactions (25 μ L each), containing 10 μ L of library DNA, 5 μ L of Q5 high fidelity buffer, 0.25 μ L of Q5 high fidelity polymerase, and 1 μ L of each Forward and Reverse common primer at 10 μ M, 0.5 μ L of 10 μ M dNTP and 7.25 μ L of pure sterile water. PCR products were pooled after amplification and size-selected (range 200 – 350 bp) using Agencourt AMPure XP magnetic beads. The size-selected library was sequenced in an Illumina NextSeq500 sequencer platform (Cardiff University Genomics Research Hub) to obtain 125bp paired-end reads.

Data processing

Paired-end reads were processed using a combination of packages in the Linux bash shell environment. First, we used GBSX v 1.3 [1] to demultiplex the paired-end reads allowing for one mismatch in the barcodes (-mb 1), no mismatch in the enzyme cut-site (-me 0), and ensuring that no common sequencing adapter was removed (-ca false). We then filtered (-qtrim r; -minlength 25) and merged the reads by individual using BBmap tools [2] mapped to *Kryptolebias marmoratus* reference genome (assembly size = 680.3Mb; number of scaffolds = 3,073; N50 = 2,229,659; GC content = 37.76%; Rhee, et al. 2017) using Bowtie 2 v. 2.2.3 with default parameters [3] and generated filtered and indexed individual BAM files with SAMtools v. 1.9 [4].

48 To call genotypes, we used ANGSD v 0.9.2.9 [5] with the following parameters:
49 minimum mapping quality (-minMapQ 30), minimum base quality (-minQ 20), only 5%
50 of missing data to consider a site or SNP (-minInd 95%), maximum of 500 reads per
51 position per individual (-setMaxDepth 500X per individual), minimum genotype
52 posterior probability (-postCutoff 0.95). Single and double-tons (SNPs that only
53 appeared in one or two individuals) were removed based on minimum minor allele
54 frequencies (-MinMaf). Anomalous reads were filtered according to the flags (-
55 remove_bads 1; SAM flag above 255). Adjusted mapping quality for excessive
56 mismatches was done using (-C 50). BAQ computation was performed (-baq 1), and the
57 minimum coverage for genotype calling was 3 (-geno_minDepth 3). We used SAMtools
58 genotype likelihood model (-GL 1) and estimated posterior genotype probabilities
59 assuming a uniform prior (-doPost 2). In addition, we used the ANGSD SNP calling
60 method (-SNP_pval 1e-6), where a Likelihood Ratio Test is used to compare between
61 the null ($\text{maf} = 0$) and alternative (estimated maf) hypotheses by using a X^2 distribution
62 with one degree of freedom.

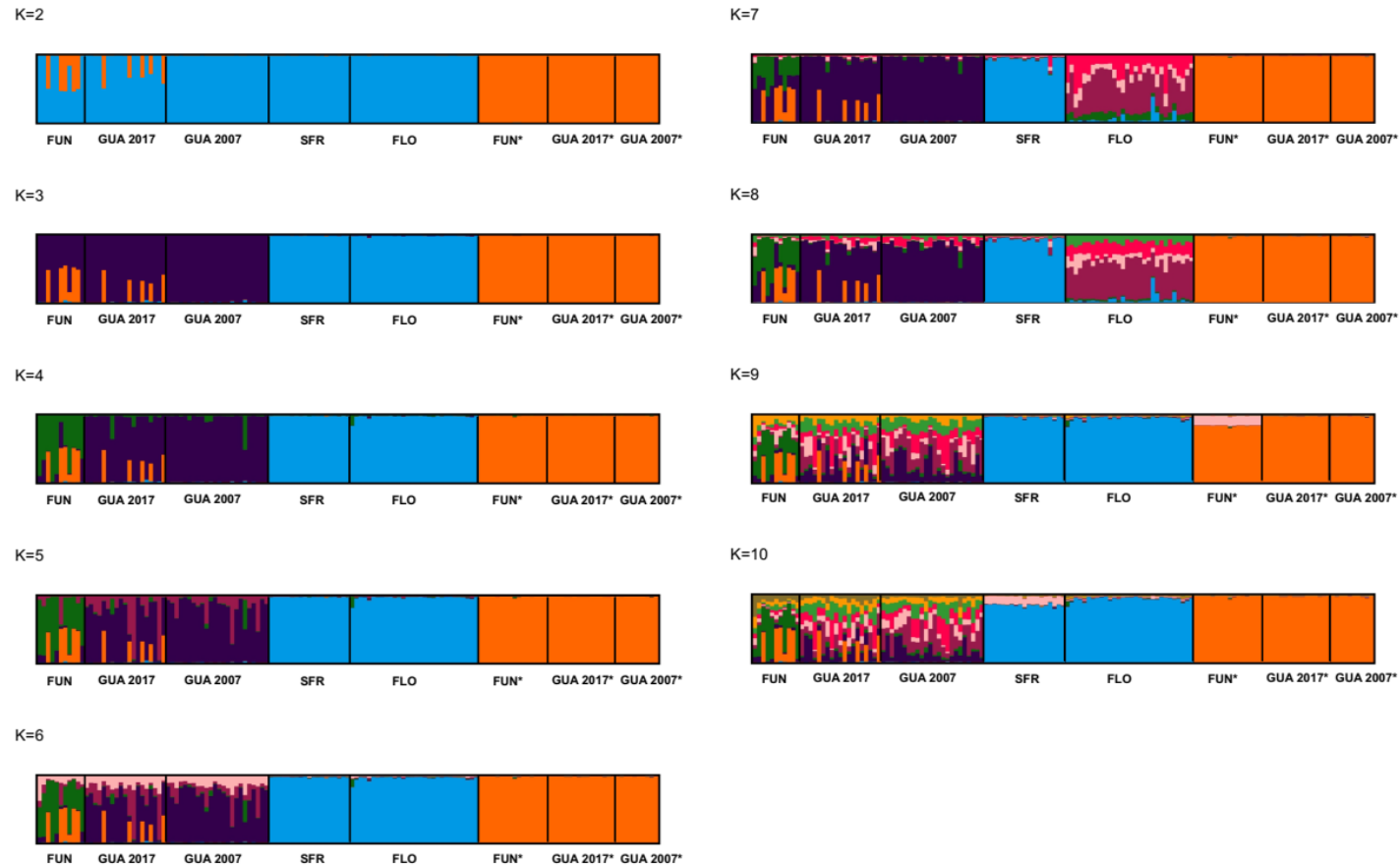


Figure S1. Individual ancestry plots with the highest likelihood runs for each K value ran in STRUCTION with genotypes for 16 microsatellites for 103 *K. ocellatus* and 42 *K. hermaphroditus* individuals. Asterisk denotes locations for individuals with *K. hermaphroditus* mtDNA.

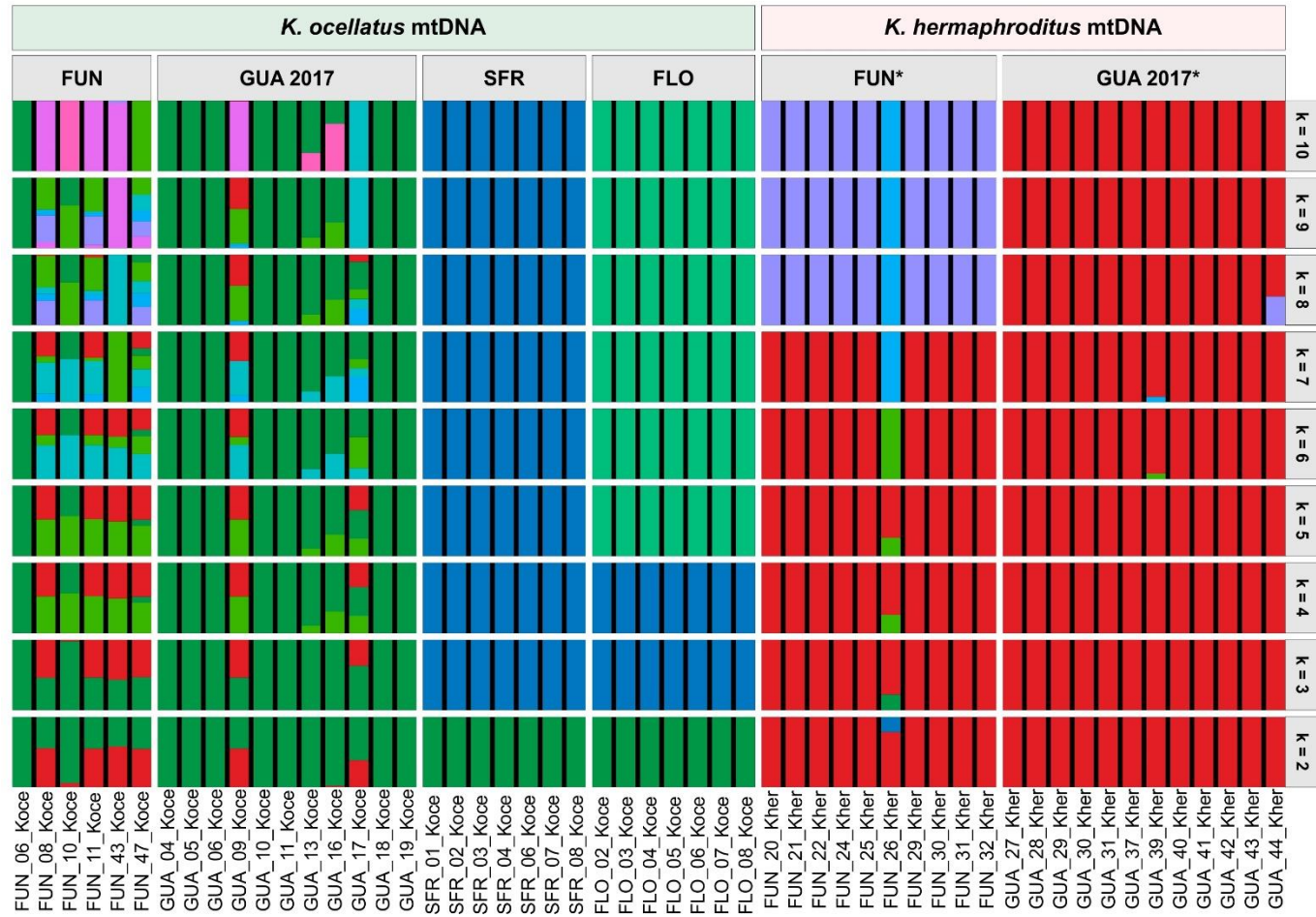


Figure S2. Individual ancestry plots with each K value ran in nsgAdmix with 5,477 SNPs from 31 *K. ocellatus* individuals and 22 *K. hermaphroditus* individuals. Asterisk denotes locations for individuals with *K. hermaphroditus* mtDNA.

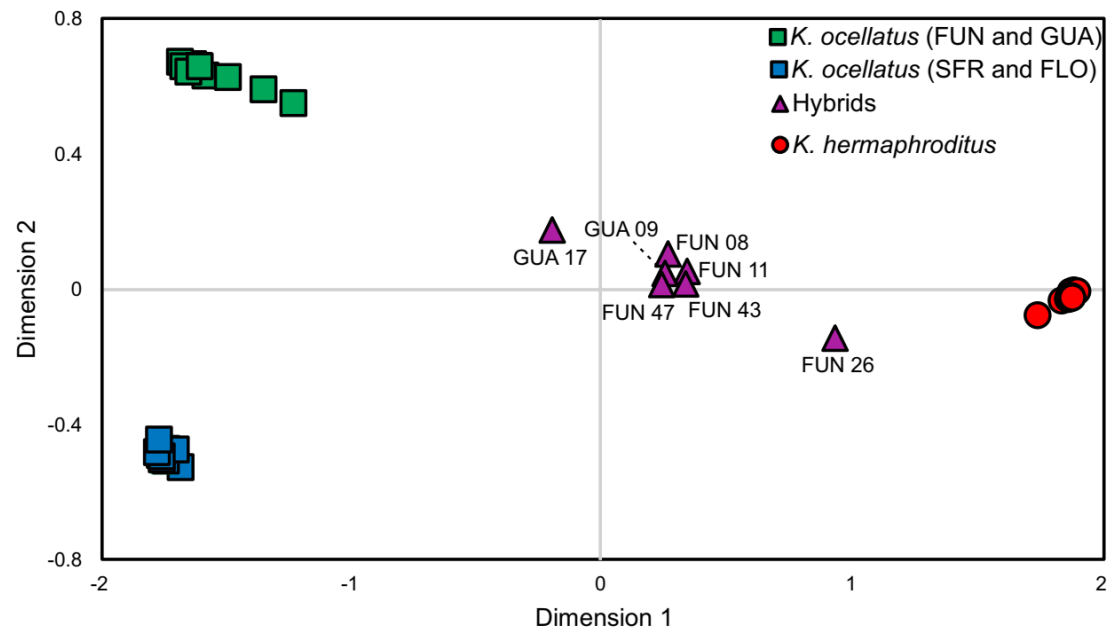


Figure S3. Multidimensional Scaling (MDS) based on the genetic distances from 5,477 SNPs. Hybrid individuals (see results) are highlighted with their respective labels.

Supplementary Table S1. Summary of 58 samples sequenced in the genotype-by-sequencing (msGBS) library. Individuals in red failed to pass the filtering cutoff (≥ 500 kb reads). Superscript letters on ID denotes F1 and backcrosses. Parameters ‘proportion of heterozygous sites’ (for Dataset I), ‘coverage and missing data’ (for Dataset II) are described in Material and Methods.

ID	Species (identified using <i>coxI</i>)	Sampling location	Number of reads	Uniquely mapped (%)	Coverage (X)	Missing data (%)	Proportion of Heterozygous sites
GUA_27	<i>K. hermaphroditus</i>	GUA	8,962,202	90.30	250.4	0.00	0.03
GUA_28	<i>K. hermaphroditus</i>	GUA	4,161,657	90.87	114.9	0.02	0.03
GUA_29	<i>K. hermaphroditus</i>	GUA	13,831,746	90.29	380.1	0.00	0.02
GUA_30	<i>K. hermaphroditus</i>	GUA	7,000,238	90.05	190.2	0.00	0.02
GUA_31	<i>K. hermaphroditus</i>	GUA	6,646,432	89.95	185.2	0.02	0.03
GUA_37	<i>K. hermaphroditus</i>	GUA	3,013,973	90.06	82.2	0.18	0.03
GUA_39	<i>K. hermaphroditus</i>	GUA	3,472,656	89.73	93.5	0.02	0.04
GUA_40	<i>K. hermaphroditus</i>	GUA	13,651,662	90.50	382.6	0.00	0.03
GUA_41	<i>K. hermaphroditus</i>	GUA	5,348,736	90.08	147.3	0.04	0.03
GUA_42	<i>K. hermaphroditus</i>	GUA	1,786,528	91.00	48.8	0.20	0.02
GUA_43	<i>K. hermaphroditus</i>	GUA	7,269,128	89.39	197.0	0.00	0.03
GUA_44	<i>K. hermaphroditus</i>	GUA	7,937,202	89.92	216.6	0.00	0.02

FUN_20	<i>K. hermaphroditus</i>	FUN	8,493,250	89.83	229.9	0.00	0.02
FUN_21	<i>K. hermaphroditus</i>	FUN	4,140,936	90.70	113.3	0.00	0.03
FUN_22	<i>K. hermaphroditus</i>	FUN	4,990,777	90.96	138.7	0.00	0.03
FUN_24	<i>K. hermaphroditus</i>	FUN	3,520,363	89.86	98.6	0.02	0.03
FUN_25	<i>K. hermaphroditus</i>	FUN	4,421,739	90.34	120.7	0.04	0.03
FUN_26 ^{Ba}	<i>K. hermaphroditus</i>	FUN	1,361,947	69.56	24.9	0.84	0.32
FUN_29	<i>K. hermaphroditus</i>	FUN	7,611,704	89.30	208.2	0.00	0.02
FUN_30	<i>K. hermaphroditus</i>	FUN	2,892,977	88.63	77.2	0.00	0.03
FUN_31	<i>K. hermaphroditus</i>	FUN	9,160,363	89.53	251.5	0.00	0.03
FUN_32	<i>K. hermaphroditus</i>	FUN	2,322,888	88.84	62.4	0.02	0.02
GUA_04	<i>K. ocellatus</i>	GUA	5,033,372	77.40	102.4	0.15	0.08
GUA_05	<i>K. ocellatus</i>	GUA	11,979,935	75.68	235.1	0.00	0.08
GUA_06	<i>K. ocellatus</i>	GUA	6,418,126	81.00	143.5	0.04	0.08
GUA_09 ^{F1}	<i>K. ocellatus</i>	GUA	5,784,849	84.54	138.1	0.05	0.61
GUA_10	<i>K. ocellatus</i>	GUA	11,097,202	81.18	243.2	0.04	0.08
GUA_11	<i>K. ocellatus</i>	GUA	2,208,760	75.05	44.5	0.46	0.06
GUA_13	<i>K. ocellatus</i>	GUA	11,446,137	77.19	230.1	0.00	0.09
GUA_16	<i>K. ocellatus</i>	GUA	14,282,577	80.18	313.2	0.00	0.11
GUA_17 ^{Ba}	<i>K. ocellatus</i>	GUA	4,522,779	72.17	91.6	0.00	0.50

GUA_18	<i>K. ocellatus</i>	GUA	7,266,866	79.34	149.5	0.05	0.08
GUA_19	<i>K. ocellatus</i>	GUA	8,150,733	81.55	177.2	0.09	0.07
GUA_20 ^{Ba}	<i>K. ocellatus</i>	GUA	20,647	NA	NA	NA	NA
GUA_24 ^{Ba}	<i>K. ocellatus</i>	GUA	4,778	NA	NA	NA	NA
GUA_62 ^{F1}	<i>K. ocellatus</i>	GUA	300,213	NA	NA	NA	NA
FUN_06	<i>K. ocellatus</i>	FUN	5,098,517	81.78	111.7	0.26	0.06
FUN_08 ^{F1}	<i>K. ocellatus</i>	FUN	9,433,950	86.75	234.2	0.00	0.60
FUN_10	<i>K. ocellatus</i>	FUN	6,856,002	81.10	145.5	0.02	0.14
FUN_11 ^{F1}	<i>K. ocellatus</i>	FUN	7,773,860	86.68	199.0	0.00	0.62
FUN_13 ^{F1}	<i>K. ocellatus</i>	FUN	2,645	NA	NA	NA	NA
FUN_41 ^{Ba}	<i>K. ocellatus</i>	FUN	7,793	NA	NA	NA	NA
FUN_43 ^{F1}	<i>K. ocellatus</i>	FUN	578,575	83.31	13.8	4.14	0.48
FUN_47 ^{F1}	<i>K. ocellatus</i>	FUN	564,546	78.82	12.4	4.16	0.50
FLO_02	<i>K. ocellatus</i>	FLO	13,605,858	81.73	301.5	0.04	0.05
FLO_03	<i>K. ocellatus</i>	FLO	2,439,264	70.34	48.0	0.38	0.06
FLO_04	<i>K. ocellatus</i>	FLO	2,167,505	79.99	46.8	0.40	0.06
FLO_05	<i>K. ocellatus</i>	FLO	2,517,261	81.12	53.6	0.66	0.06
FLO_06	<i>K. ocellatus</i>	FLO	4,609,454	78.83	96.2	0.11	0.06
FLO_07	<i>K. ocellatus</i>	FLO	4,620,891	81.46	96.3	0.02	0.05

FLO_08	<i>K. ocellatus</i>	FLO	6,823,947	79.43	149.9	0.00	0.06
SFR_01	<i>K. ocellatus</i>	SFR	9,375,699	81.74	209.4	0.05	0.06
SFR_02	<i>K. ocellatus</i>	SFR	2,671,245	82.54	60.3	0.37	0.06
SFR_03	<i>K. ocellatus</i>	SFR	8,014,672	81.64	177.9	0.11	0.06
SFR_04	<i>K. ocellatus</i>	SFR	8,136,965	83.09	177.8	0.00	0.07
SFR_06	<i>K. ocellatus</i>	SFR	10,875,988	82.37	240.8	0.15	0.05
SFR_07	<i>K. ocellatus</i>	SFR	10,572775	82.07	235.7	0.05	0.06
SFR_08	<i>K. ocellatus</i>	SFR	684,671	77.07	14.5	4.93	0.05
Average			6,370,039	83.90	152.98	0.34	0.11

Supplementary Table S2. Descriptive statistics of genetic variation at microsatellite loci in (a) 103 *Kryptolebias ocellatus* and (b) 42 *K. hermaphroditus* individuals. N = sample size; P₉₉ = proportion of polymorphic loci (99% criterion); A = average number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity. Sampling locations are described in Table 1.

Sampling location	N	P ₉₉	A	H _E	H _O
(a) Individuals with <i>K. ocellatus</i> mtDNA					
GUA 2007	24	0.93	8.19	0.60	0.53
GUA 2017	19	1.00	9.31	0.65	0.64
FUN 2017	11	0.93	6.56	0.66	0.70
SFR	19	0.87	7.25	0.57	0.63
FLO	30	0.81	9.88	0.54	0.52
Mean	20.6	0.90	3.83	0.60	0.60
(b) Individuals with <i>K. hermaphroditus</i> mtDNA					
GUA 2007	10	0.21	1.57	0.10	0.00
GUA 2017	16	0.13	1.73	0.09	0.01
FUN 2017	16	0.13	1.33	0.03	0.00
Mean	14	0.15	1.54	0.07	0.003

References

1. Herten, K.; Hestand, M.S.; Vermeesch, J.R.; Van Houdt, J.K. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* **2015**, *16*, 73, doi:10.1186/s12859-015-0514-3.
2. Bushnell, B. *BBMap: a fast, accurate, splice-aware aligner*; Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States): 2014.
3. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **2012**, *9*, 357.
4. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078-2079.
5. Korneliussen, T.S.; Albrechtsen, A.; Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **2014**, *15*, 356, doi:10.1186/s12859-014-0356-4.