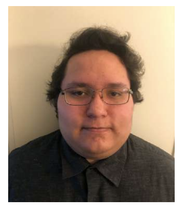
Copy Number Variation and Divergence in Color and Morphology in a Threespine Stickleback Stream-Anadromous Pair

Grant Tiger, Bhakti Vahewala (presenting), Lucas Cortés and Jeffrey S. McKinnon[1],
Dept. Biology, East Carolina U., Greenville, NC, USA **[1]mckinnonj@ecu.edu**

Background

- The threespine stickleback is a teleost that has undergone recurrent adaptive radiations from marine to freshwater environments.
- The Little Campbell stream-anadromous pair is an example of recently diverged, but closely related, morphs. One way in which they are divergent is in their red throat coloration. Females in the stream population often have characteristic red throats, whereas anadromous females lack this trait.
- Diploid organisms typically carry two copies of each gene. However, differences in the number of copies of chromosomal segments exist in the form of Copy Number Variation (CNV) which has arisen as a key feature of evolutionary diversification.

Figure 1: Threespine stickleback from Little Campbell Stream population. Note characteristic red throat. (Credit: Christopher Anderson)

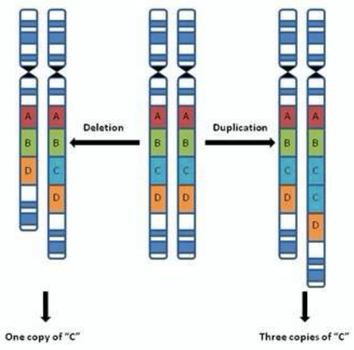


Figure 2: Copy number variation is the change in copy number of chromosomal segments. The chromosomal segment C appears on each chromosome in a normal copy number, it only appears once in a heterozygous deletion (left) and appears three times in a heterozygous duplication (right) (Neurowiki, 2013).

Central Aims of Study:

- 1) Identify and quantify the CNV differences between the two study populations (anadromous/stream)
- 2) Identify and quantify the CNV differences between red and dull stickleback females
- 3) Identify and quantify CNV in specific color genes previously identified in a stickleback gene expression study (Newsome *et al*. unpublished)

Figure 3: Little Campbell Stream, British Columbia, Canada (source: google.com/earth)

Methods

- Collected fish, Little Campbell Stream (LCS) and Little Campbell Anadromous (LCA), from British Columbia
- Extracted DNA from fin clips
- Sequenced DNA using Illumina paired-end sequencing (30x depth), aligned reads to the reference stickleback genome
- GATK CNVcaller was used to identify deletions and duplications in relation to the reference genome
- Statistically verified CNV differences between populations and color morphs. Identified CNVs in 18 specific genes, previously identified in Newsome *et al.* unpublished by scanning VCF files

```
GATK.sh
../../programs/gatk-4.1.2.0/gatk-package-4.1.2.0-local.jar AnnotateIntervals
-R revisedAssemblyUnmasked.fa
-L preprocessed_intervals.interval_list
--interval-merging-rule OVERLAPPING_ONLY
-O annotated_intervals.tsv

../../programs/gatk-4.1.2.0/gatk-package-4.1.2.0-local.jar FilterIntervals
-L preprocessed_intervals.interval_list
--annotated-intervals annotated_intervals.tsv
--interval-merging-rule OVERLAPPING_ONLY
-O filtered_intervals.interval_list

../../programs/gatk-4.1.2.0/gatk-package-4.1.2.0-local.jar CollectReadCounts
-I 5270-S2_S2_trimmedaligned_unique_fixed.bam.sorted1.bam
-L filtered_intervals.interval_list
--interval-merging-rule OVERLAPPING_ONLY
-O 5270-S2_S2.counts.hdf5

for i in *.sorted.bam;
 do code="echo ${i} | cut -d'_' -f1,2";
 ../../programs/gatk-4.1.2.0/gatk-package-4.1.2.0-local.jar CollectReadCounts
 -I ${i}
 -L filtered_intervals.interval_list
 --interval-merging-rule OVERLAPPING_ONLY
 -O $code.counts.hdf5; done
```

Figure 4: Code for separating BAM files into intervals which are later used for CNV calling

```
##source=PostprocessGermlineCNVcalls
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  5270-S32_S32
chrI    1       CNV_chrI_1_1000 N       <DEL>,<DUP>     .       .       END=1000        GT:CN:CNLP:CNQ  0:2:309,62,0
chrI    1001    CNV_chrI_1001_2000      N       <DEL>,<DUP>     .       .       END=2000        GT:CN:CNLP:CNQ  0:2:
chrI    2001    CNV_chrI_2001_3000      N       <DEL>,<DUP>     .       .       END=3000        GT:CN:CNLP:CNQ  0:2:
chrI    3001    CNV_chrI_3001_4000      N       <DEL>,<DUP>     .       .       END=4000        GT:CN:CNLP:CNQ  0:2:
chrI    4001    CNV_chrI_4001_5000      N       <DEL>,<DUP>     .       .       END=5000        GT:CN:CNLP:CNQ  0:2:
chrI    5001    CNV_chrI_5001_6000      N       <DEL>,<DUP>     .       .       END=6000        GT:CN:CNLP:CNQ  0:2:
chrI    6001    CNV_chrI_6001_7000      N       <DEL>,<DUP>     .       .       END=7000        GT:CN:CNLP:CNQ  0:2:
chrI    7001    CNV_chrI_7001_8000      N       <DEL>,<DUP>     .       .       END=8000        GT:CN:CNLP:CNQ  0:2:
chrI    8001    CNV_chrI_8001_9000      N       <DEL>,<DUP>     .       .       END=9000        GT:CN:CNLP:CNQ  0:2:
chrI    9001    CNV_chrI_9001_10000     N       <DEL>,<DUP>     .       .       END=10000       GT:CN:CNLP:CNQ  0:2:
chrI    10001   CNV_chrI_10001_11000    N       <DEL>,<DUP>     .       .       END=11000       GT:CN:CNLP:CNQ  0:2:
chrI    11001   CNV_chrI_11001_12000    N       <DEL>,<DUP>     .       .       END=12000       GT:CN:CNLP:CNQ  0:2:
chrI    12001   CNV_chrI_12001_13000    N       <DEL>,<DUP>     .       .       END=13000       GT:CN:CNLP:CNQ  0:2:
chrI    13001   CNV_chrI_13001_14000    N       <DEL>,<DUP>     .       .       END=14000       GT:CN:CNLP:CNQ  0:2:
chrI    17001   CNV_chrI_17001_18000    N       <DEL>,<DUP>     .       .       END=18000       GT:CN:CNLP:CNQ  0:2:
chrI    18001   CNV_chrI_18001_19000    N       <DEL>,<DUP>     .       .       END=19000       GT:CN:CNLP:CNQ  0:2:
chrI    19001   CNV_chrI_19001_20000    N       <DEL>,<DUP>     .       .       END=20000       GT:CN:CNLP:CNQ  0:2:
chrI    20001   CNV_chrI_20001_21000    N       <DEL>,<DUP>     .       .       END=21000       GT:CN:CNLP:CNQ  0:2:
chrI    21001   CNV_chrI_21001_22000    N       <DEL>,<DUP>     .       .       END=22000       GT:CN:CNLP:CNQ  0:2:
chrI    22001   CNV_chrI_22001_23000    N       <DEL>,<DUP>     .       .       END=23000       GT:CN:CNLP:CNQ  0:2:
chrI    23001   CNV_chrI_23001_24000    N       <DEL>,<DUP>     .       .       END=24000       GT:CN:CNLP:CNQ  0:2:
chrI    24001   CNV_chrI_24001_25000    N       <DEL>,<DUP>     .       .       END=25000       GT:CN:CNLP:CNQ  0:2:
```

Figure 5: VCF file containing the intervals in the genome that have a CNV within them (credit: Grant Tiger)

Results, Interpretation

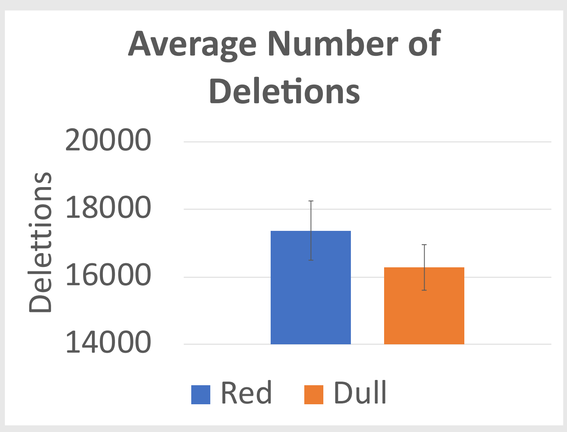


***Figure 6*:** No significant difference in deletions between red and dull females

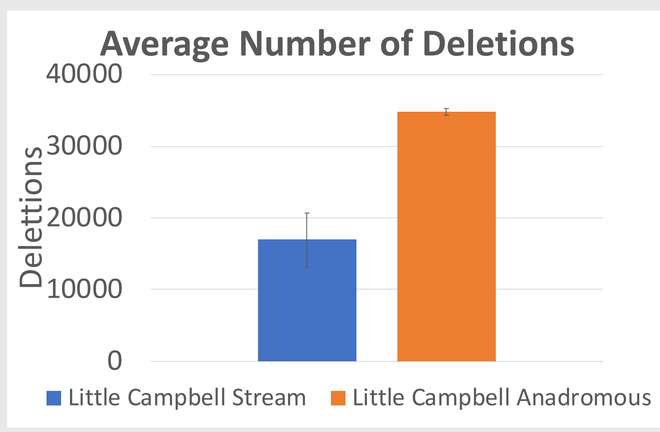


***Figure 8*:** Significantly more deletions in Little Campbell Anadromous females than in Little Campbell Stream females

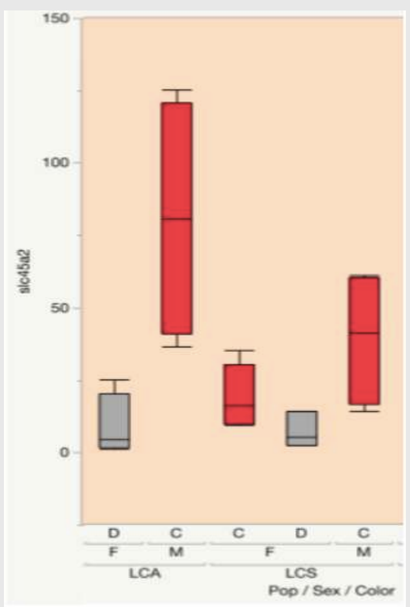


Figure 10: Expression patterns of LCS and LCA populations (D=dull, C=colored, F=female, M=male)

***Table 1*:** Deletions in gene *slc45a2* are restricted to only anadromous individuals

ENSGACG00000016297				
Name	Sample Number	Position	Copy Number State	Number of Copies
LCAF1	90	3523001	1	0
LCAF2	91	3523001	1	0
LCAF3	92	3523001	1	0
LCAF4	93	3523001	1	0
LCAF5	94	3523001	1	0
AFT2	36	3523001	1	0
AFT4	37	3523001	1	0
		3524001	1	0
AFT5	38	3523001	1	0

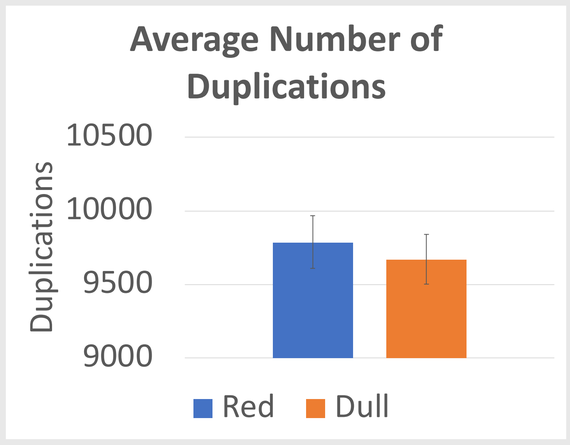


Figure 7: No significant difference in duplications between red and dull females

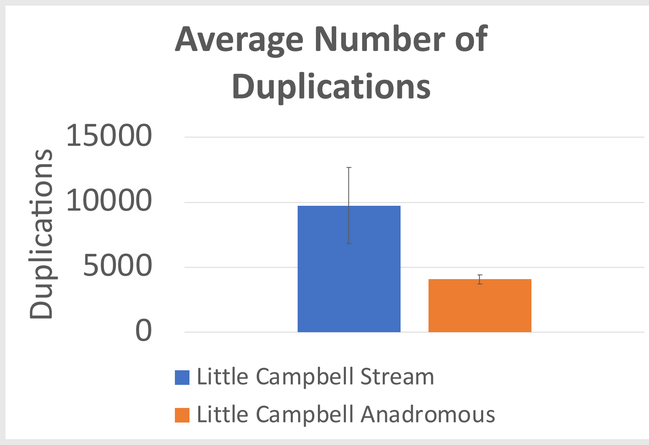


***Figure 9*:** Significantly more duplications in Little Campbell Steam females than in Little Campbell Anadromous females

Conclusions

- Stream individuals show lower levels of deletions indicating that they can act as reservoirs for ancient alleles selected out of marine populations (Lowe *et al*., 2018)
- Freshwater (stream) individuals generally have more gene copies than anadromous, possibly providing new gene functions for freshwater invasion (Hirase *et al.*, 2014)
- There were no CNV differences found between the red and dull females
- We found very few CNVs within genes. This is expected as it is thought that around 95% of CNVs are in noncoding regions (Lowe *et al*., 2018)
- Gene *slc45a2*, which is responsible for melanocyte differentiation in zebrafish, was the only gene to contain CNVs (deletions) in the coding region, compared to the reference genome. The CNVs were exclusively found in the anadromous population which suggests that the deletions could be responsible for the lack of female color in this population.

Future Direction

- Hope to search more thoroughly for CNVs in intronic regions near genes of interest
- Will include more study populations to enhance and assess generality
- Use additional calling programs and molecular technique to validate calls

Acknowledgements: NIH

References: Hirase, S., Ozaki, H. and Iwasaki, W. (2014). Parallel selection on gene copy number variations through evolution of three-spined stickleback genomes. *BMC Genomics*, 15(1), p.735. Lowe, C.B., Sanchez-Luege, N., Howes, T.R., Brady, S.D., Daugherty, R.R., Jones, F.C., Bell, M.A. and Kingsley, D.M. (2018). Corrigendum: Detecting differential copy number variation between groups of samples. *Genome Research*, 28(5), pp.766.1-766.1. Neurowiki2013.wikidot.com. 2013. *Copy Number Variations - Neurowiki 2013*. [online] Available at: <http://neurowiki2013.wikidot.com/individual:copy-number-variations> [Accessed 1 April 2020].