A pipeline for analysis of allele specific expression from RNA-seq data reveals salinity-dependent response in Nile tilapia

Aurora Campo[1], Moran Gershoni[1], Adi Doron-Faigenboim[1], and Avner Cnaani[1]

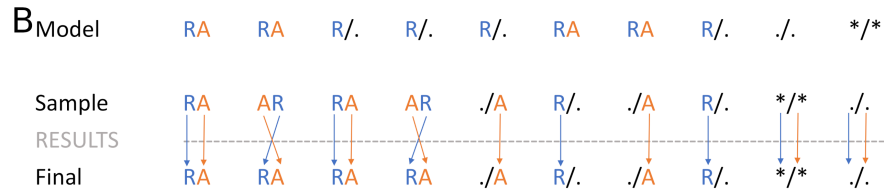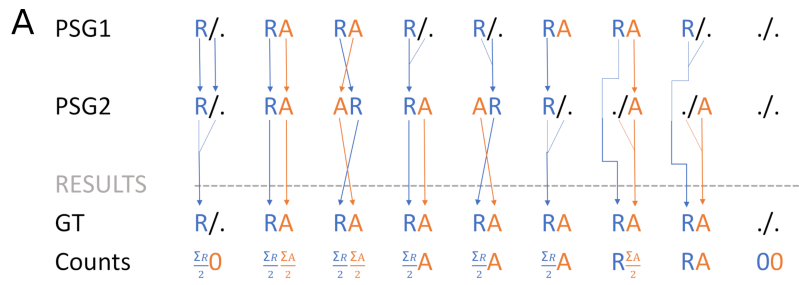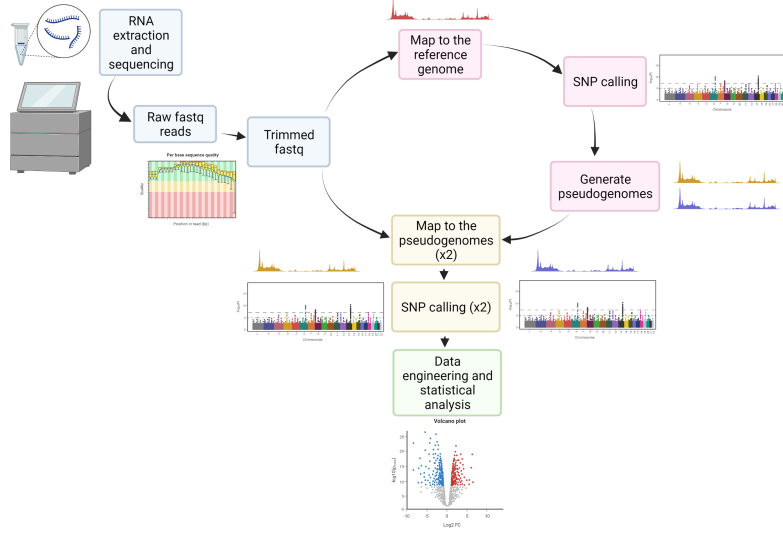[1]Agricultural Research Organization Volcani Center

June 1, 2022

Abstract

Species living in a changing environment are capable of adapting to alterations of various factors. Physiological acclimatization may be significantly influenced by the heterozygosity, especially with regards to allele variance and its specific expression (ASE) under different conditions. Data from RNA-seq experiments can be used to identify and quantify the alleles expressed, in order to detect and characterize ASE and regulation of gene expression. However, the allele matching the reference genome creates a mapping bias that prevents a reliable estimation of the allele depth unless the haplotype of the experimental individuals is provided. We developed a pipeline that allows the identification of the alleles corresponding to an RNA-seq dataset and their unbiased quantification. This pipeline does not require the sequencing of the DNA nor the previous knowledge of the haplotype. The identified SNPs are further substituted in the reference genome, thus creating two pseudogenomes with the alternative alleles on two independent samples of the experiment. The SNPs are further called against each pseudogenome thus providing with two SNP datasets that are averaged for calculation of the allele depth. The final SNP calling file contains the coordinates of the SNPs and also the ID of genes containing the SNPs, the expressed genotypes, the unbiased allele depth and the statistical tests for identifying ASE according to the experimental design and correlated with differentially expressed genes. Therefore, the pipeline presented here can calculate ASE in non-model organisms and can be applied to previous RNA-seq datasets for expanding studies in gene expression regulation.
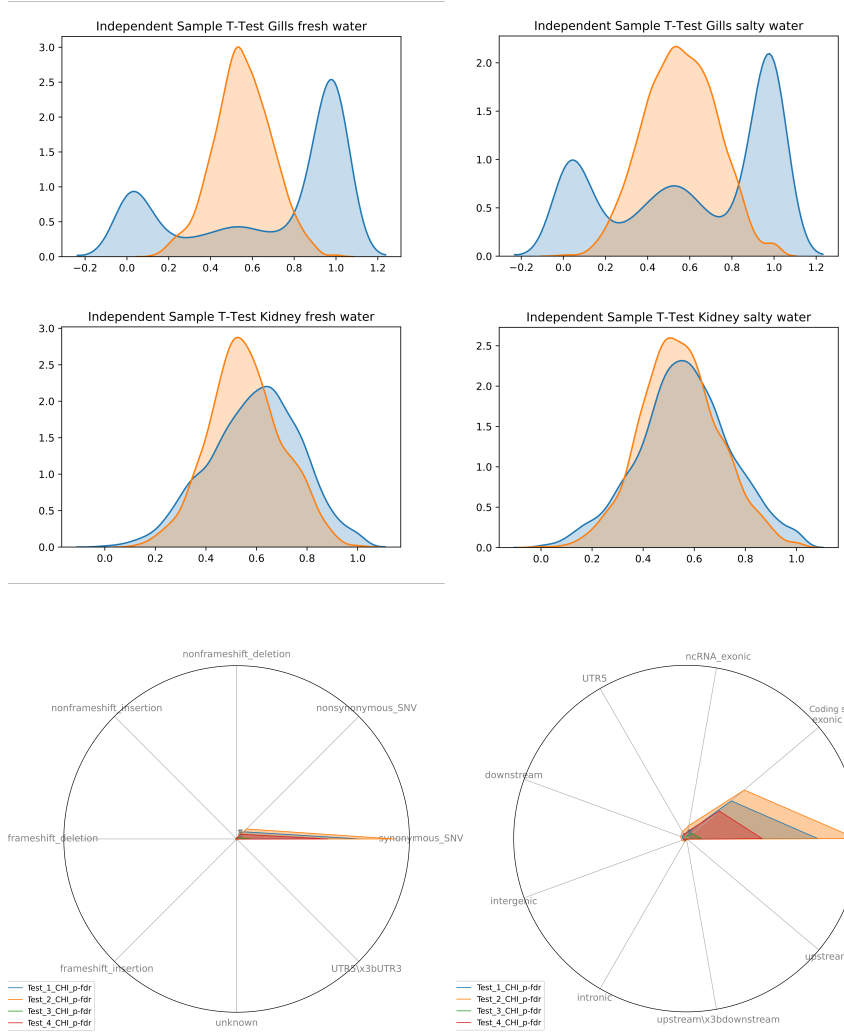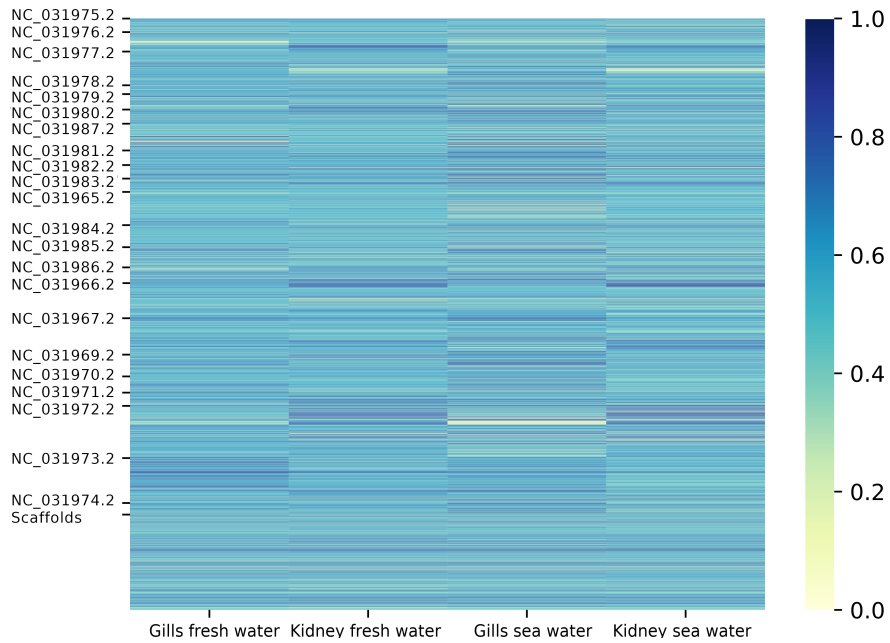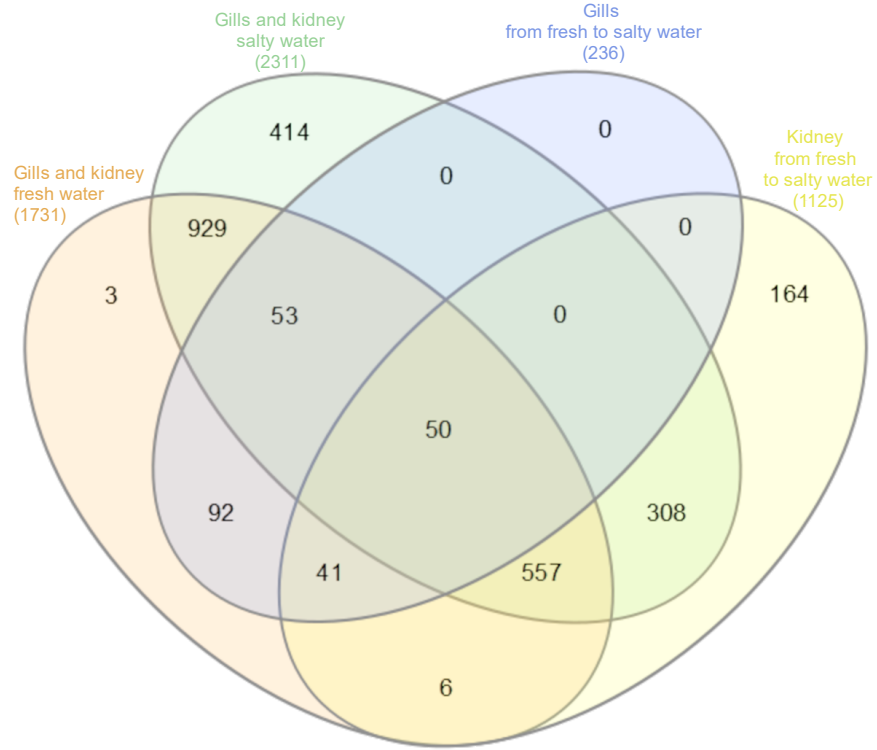
Hosted file

ASE pipeline.pdf available at https://authorea.com/users/376225/articles/571379-a-pipeline-for-analysis-of-allele-specific-expression-from-rna-seq-data-reveals-salinity-dependent-response-in-nile-tilapia

1

A

	PSG1	R/.	RA	RA	R/.	R/.	RA	RA	R/.	./.
	PSG2	R/.	RA	AR	RA	AR	R/.	./A	./A	./.
	RESULTS									
	GT	R/.	RA	RA	RA	RA	RA	RA	RA	./.
	Counts	$\frac{\Sigma R}{2}$0	$\frac{\Sigma R}{2}\frac{\Sigma A}{2}$	$\frac{\Sigma R}{2}\frac{\Sigma A}{2}$	$\frac{\Sigma R}{2}$A	$\frac{\Sigma R}{2}$A	$\frac{\Sigma R}{2}$A	$\frac{\Sigma A}{2}$	RA	00

B

	Model	RA	RA	R/.	R/.	R/.	RA	RA	R/.	./.	*/*
	Sample	RA	AR	RA	AR	./A	R/.	./A	R/.	*/*	./.
	RESULTS										
	Final	RA	RA	RA	RA	./A	R/.	./A	R/.	*/*	./.

/ : Model in the next comparison

./. : Not expressed in this sample

Independent Sample T-Test Gills fresh water

Independent Sample T-Test Gills salty water

Independent Sample T-Test Kidney fresh water

Independent Sample T-Test Kidney salty water

Gills and kidney in fresh water
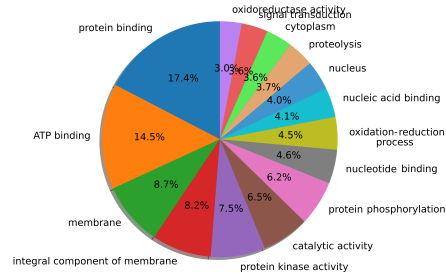
Gills and kidney in salty water

Salinity in gills

Salinity in kidney

Gills and kidney in fresh water

Gills and kidney salty water