**Appendix S3**. Perl script (*Hit-desc.pl*) for adding functional annotations to contigs using BLAST output (-outfmt 6). *Note*: This script uses regular expressions that are specific to targeting the accession numbers of our contigs and those that were used for BLAST searches. It is to be used as a reference for writing similar scripts for similar purposes but will not work as is for data sets other than our own. Requires an installation of BioPerl. Running too much BLAST output at once will kick you off of the NCBI server, so restrict the number of queries to ~500.

Run at the command line as:

perl Hit-desc.pl BLASToutput.txt outfile.txt

#!/usr/bin/perl -w

use strict;

use Bio::DB::GenBank;

######################################################################

## This program is designed to retrieve accession numbers for a ####

## query sequence and a hit sequence, and will display the name ####

## and the description of the hit sequence. ####

######################################################################

## Open appropriate files. ##

open INFILE, '<', $ARGV[0];

open HITS, '>', $ARGV[1];

my ($query\_seq, $hit\_seq);

## Text manipulations to get accession numbers. ##

my @lines = <INFILE>;

my $gb = Bio::DB::GenBank->new();

foreach (@lines){

if(/(AKK\w\d+.\d)/){

$query\_seq=$1;

}

if(/((N|Y)P\_\d+.\d)/){

$hit\_seq=$1;

}

my $seq = $gb->get\_Seq\_by\_version($hit\_seq);

print HITS "$query\_seq\t$hit\_seq\t", $seq->desc, "\n";

}