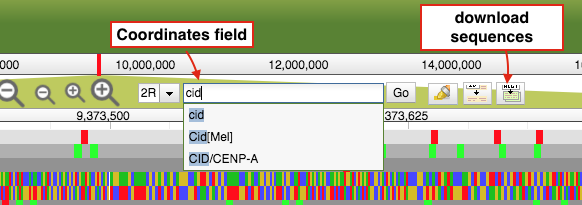
Original doc by Lisa Kursel. Updates by Janet Young

1. Go to http://popfly.uab.cat/

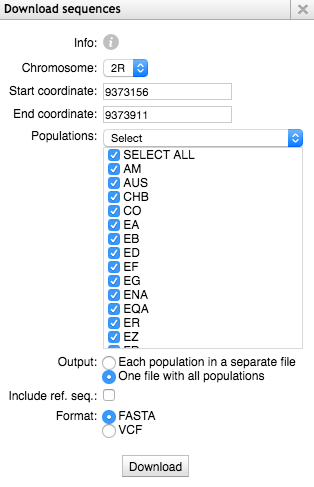
2. Enter gene name into coordinates field and hit 'Go'



3. Highlight your gene so it fills the entire viewing window (you’ll have to trim UTRs and introns later)

4. Click the *download sequences* button (see picture in step 2). A little window will pop-up…

5. Under the *populations* menus, choose *ZI* (it’s better to use a single population, and this one has lots of strains sampled). Then click *Download*. The download might take a few minutes.



6. Unzip the file downloaded FASTA file. Import the FASTA file as an alignment into your sequence editing software of choice. There may be duplicates in this alignment (sequences that have the same name). These will be removed later. Don’t change their names.

7. Trim your sequence alignment so that it is only coding sequence (remove UTRs, introns). Save this trimmed alignment as a FASTA file.

8. Run a perl script to remove all sequences that contain Ns.

removeSeqsContainingNs.pl GeneOfInterestAlignment\_CDS.fasta

(you will see a new file appear – its name will end in noNseqs.fa)

9. Get D. simulans ortholog, and make an in-frame alignment with your D.mel sequences

10. Run MK test! Use this website: <http://mkt.uab.es/mkt/mkt.asp>

1. Paste your D. mel sequences into the ‘species 1’ box, and the D.sim sequence into the ‘species 2’ box.
2. Click on the ‘main parameters’ tab, and unselect ‘align sequences’ (it is very slow! better to align it yourself).
3. It’s also recommended that you select ‘exclude low frequency variants’ using a 5% threshold (also under ‘main parameters’
4. Click ‘Run test’.
5. You’ll probably want to see the results ‘without any correction for divergence’