**User Guide for the LRG Parser *lrgparser.py***

Developers: Rebecca Haines, Laura Lopez-Pascua, Sara Rey and Ben Sanders

Contact: ben.sanders-2@postgrad.manchester.ac.uk

Release date: 14th November 2014

Program name: lrgparser.py

The program *lrgparser.py* takes a Locus Reference Genomic file in xml format downloaded from http://www.lrg-sequence.org/ and outputs a FASTA formatted file that contains the sequences for each exon in the genomic sequence. This version of the tool supports version 1.8 LRG files with a single transcript. LRG files with multiple transcripts are not supported.

System requirements:

This program was written in Python. It is expected that is will be run from the terminal on a computer running Linux. It has not been tested in any other environments.

Instructions for use:

1. Go to menu -> accessories -> LXTerminal
2. Open the folder where the program is stored. Eg if it is stored in a folder called “programs” in the home directory type cd programs
3. Check that your LRG file is in the same folder as the program
4. To run to program type:

python lrgparser.py <LRGFileName>

1. The FASTA file will be saved in the same folder, with the same name as the LRG file, but with the file extension .fa

Troubleshooting:

If an error appears check the text of the error. Some suggested solutions are listed below.

Error: Usage: python lrgparser.py <input\_file>

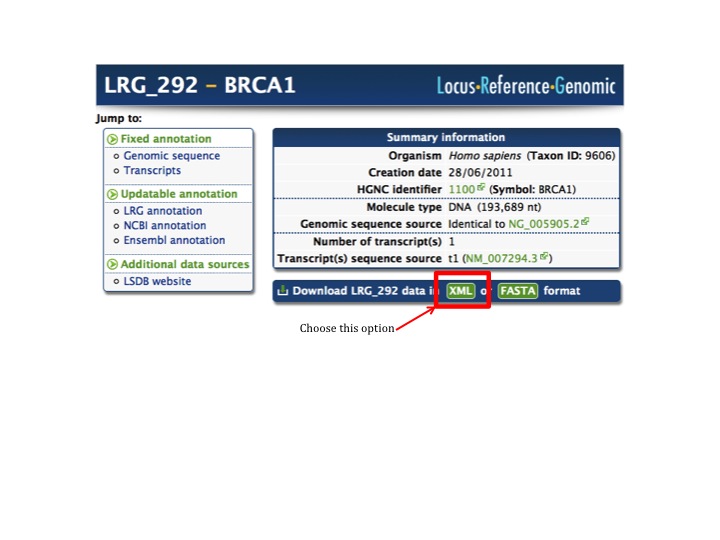
Check that you have typed the program running command, including the name of the file, eg python lrgparser.py LRG\_292.xml

Error: Incorrect schema version. This tool only supports version 1.8

The program lrgparser.py has been designed to work with version 1.8 LRG files. If the file you wish to work with is a different version then this program may not work. Please contact the developer.

Error: Not an xml file or ElementTree not installed

Check that you have downloaded the .xml format file from lrg-sequence.org, rather than the FASTA file (see below).

**

Error:

* The pulled sequence's length is 0
* Mistake- sequence contains characters other than A, T, C or G (note: case sensitive)
* input coordinates are not integers
* List of exon co-ordinates is empty. No exon co-ordinates stored.
* Error: Missing co-ordinate in list of exon co-ordinates
* Second co-ordinate of co-ordinate pair bigger than first
* Exon n starts before the end of exon n-1 or this LRG file may contain alternate transcripts
* Co-ordinates lie outside of the range of the sequence
* Exon is not long enough, it has fewer than 1 nucleotide
* Accession number has no assigned value
* LRG ID has no assigned value
* Gene name has no assigned value

All of these errors indicate that there is an issue with the way the program has read the LRG file. It will not be possible to use the lrgparser.py program with this file.

Error: could not open output file

The program has failed to create the FASTA file. Contact the developer for advice.