

Documentation

Table 1: General information about the suite of software to finish genome assemblies

Available at:	http://github.com/dcbmariano/scripts
General requirements:	<ul style="list-style-type: none">Operational system: Linux 64bitPerlPythonBiopython library<ul style="list-style-type: none">http://biopython.orgNCBI-BLAST+<ul style="list-style-type: none">http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=DownloadMira 4.0 Assembler<ul style="list-style-type: none">http://mira-assembler.sourceforge.net
Scripts:	<ul style="list-style-type: none">CONTIGuatorD.pycontiginfo.pycut_left.plmapRepeat.pymcontig.pymoveDNAA.py
Advanced install:	<p>Paste the scripts on the folder ~/bin/ Open the folder on terminal and give permission to running: <code>sudo chmod +x *</code> Now you can run the scripts of two ways:</p> <ol style="list-style-type: none">Typing “python ~/bin/script_name.py parameter_1 parameter_n”Typing just “script_name parameter_1 parameter_n”

Table 2: Description of the Scripts

Script	Language	Function	Syntax
CONTIGuatorD.py	Python	Scaffolding contigs using a reference genome	python CONTIGuatorD.py -c [contigs_fasta_file] -r [reference_fasta_file] -g [reference_genbank_file]
contiginfo.py	Python	Analyze fasta files. It generates information about: number of contigs, min contig, max contig, length of genome and N50 value	python contiginfo.py [fasta_file]
cut_left.pl	Perl	Allow performing cuts on the left region of fasta or fastq files	perl cut_left.pl [len_cut] [fasta_or_fastq_file]
mapRepeat.py	Python	Allow the scaffolding of the contigs obtained by <i>de novo</i> assembly, including repetitive	python mapRepeat.py [contigs aligned file] [reference file] [fastq xml folder] [contig left name]

		regions based on the extraction of the consensus sequence from the reads mapped into the reference genome	[contig right name]
mcontig.py	Python	Create fasta sequences separating N's regions.	python mcontig.py [fasta_file]
moveDNAA.py	Python	Correct the beginning of circular genomes finding the gene dnaA by a reference fasta file	python movednaa.py [seq] [reference fasta file]