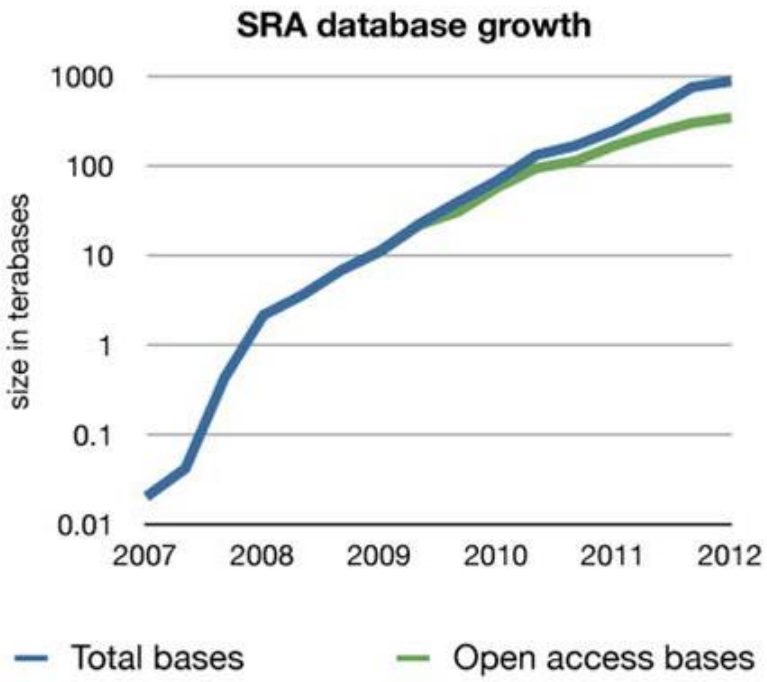
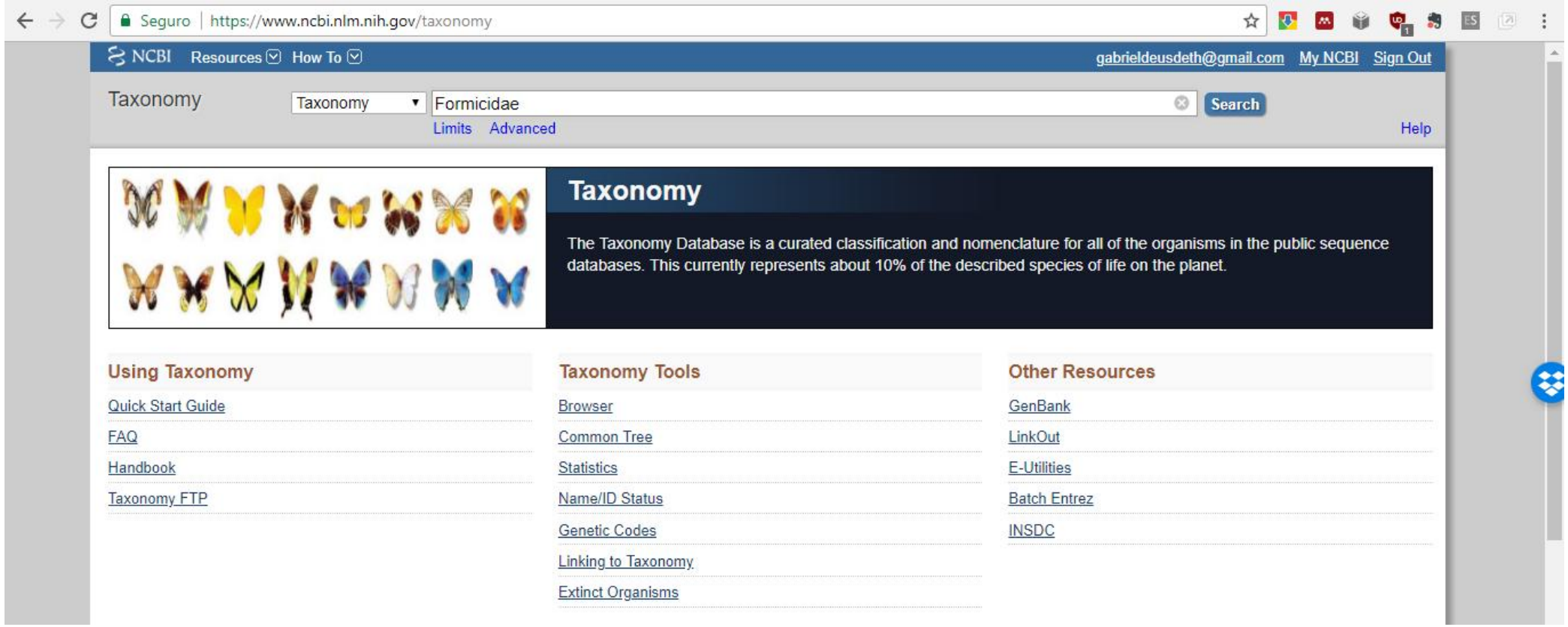
TUTORIAL: MONTAGEM DE GENOMAS USANDO DADOS PUBLICOS

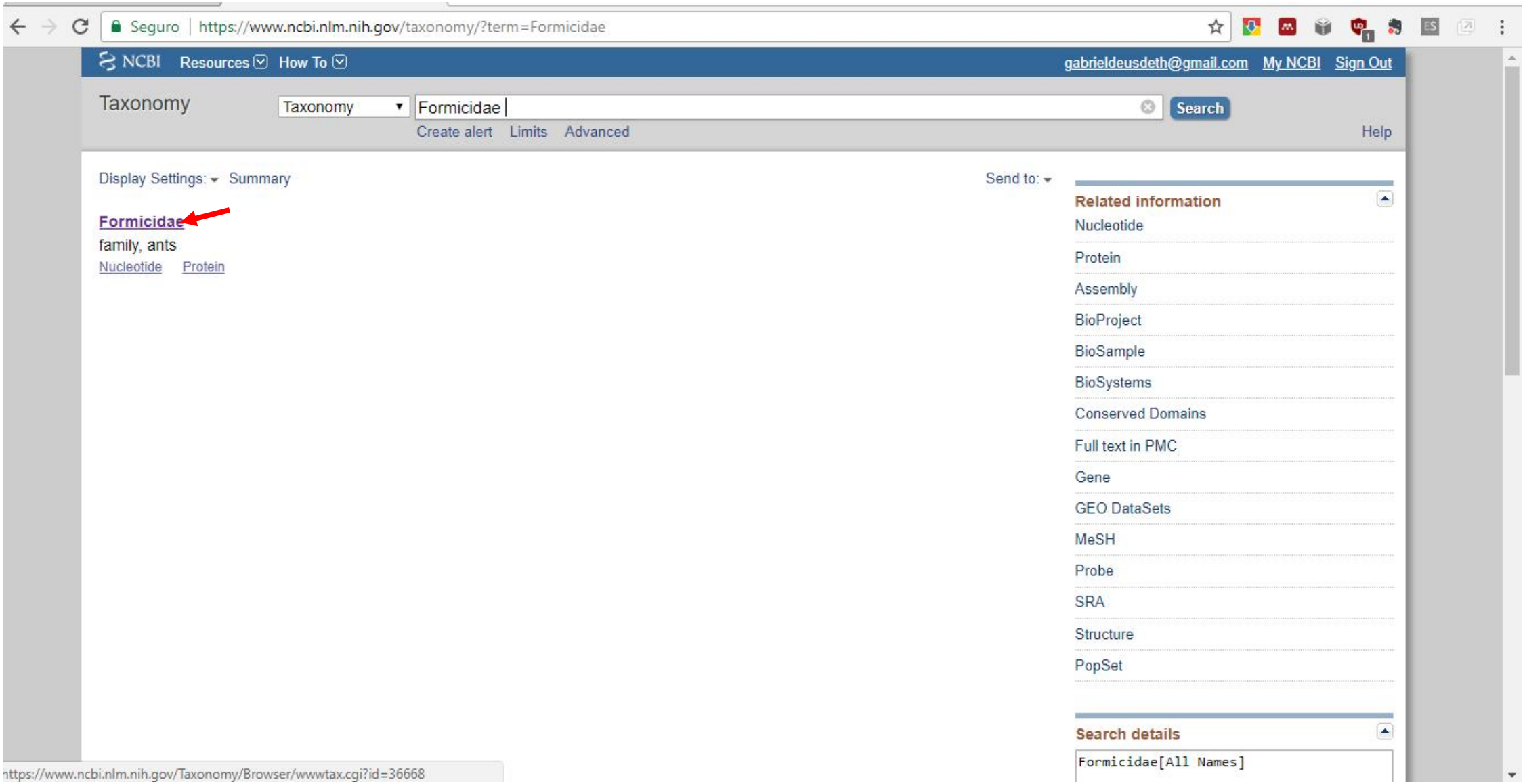


Escolhendo seu Pokémon: NCBI taxonomy

- Entre no NCBI Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy)
- Digita o nome do grupo de interesse e clique em “search”



- Na próxima página, clique no grupo desejado



- Na próxima página, aparecerão todos os resultados

NCBI Taxonomy Browser

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search for as complete name lock Go

Display 3 levels using filter: none

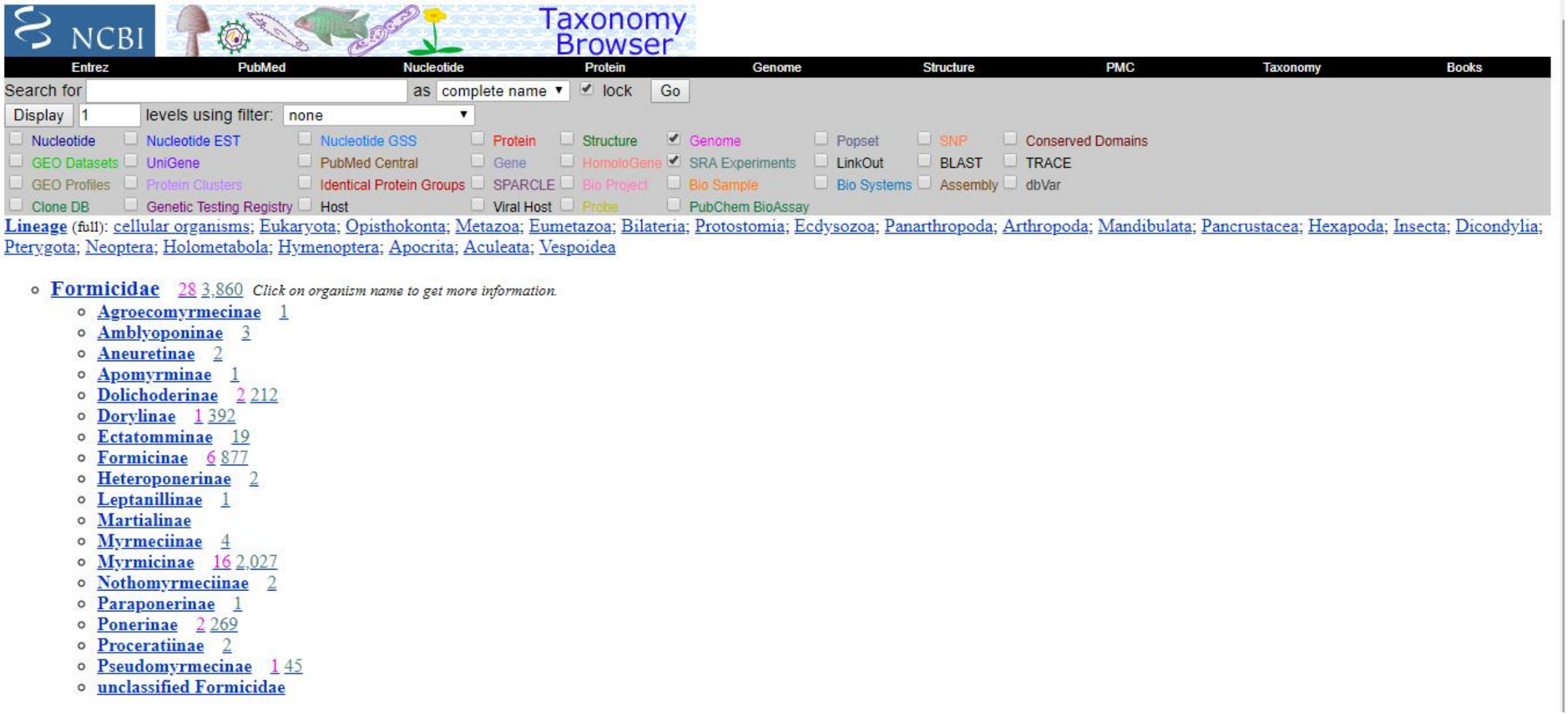
Nucleotide, Nucleotide EST, Nucleotide GSS, Protein, Structure, Genome, Popset, SNP, Conserved Domains
GEO Datasets, UniGene, PubMed Central, Gene, HomoloGene, SRA Experiments, LinkOut, BLAST, TRACE
GEO Profiles, Protein Clusters, Identical Protein Groups, SPARCLE, Bio Project, Bio Sample, Bio Systems, Assembly, dbVar
Clone DB, Genetic Testing Registry, Host, Viral Host, Probe, PubChem BioAssay

Lineage (full): cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Holometabola; Hymenoptera; Apocrita; Aculeata; Vespoidea

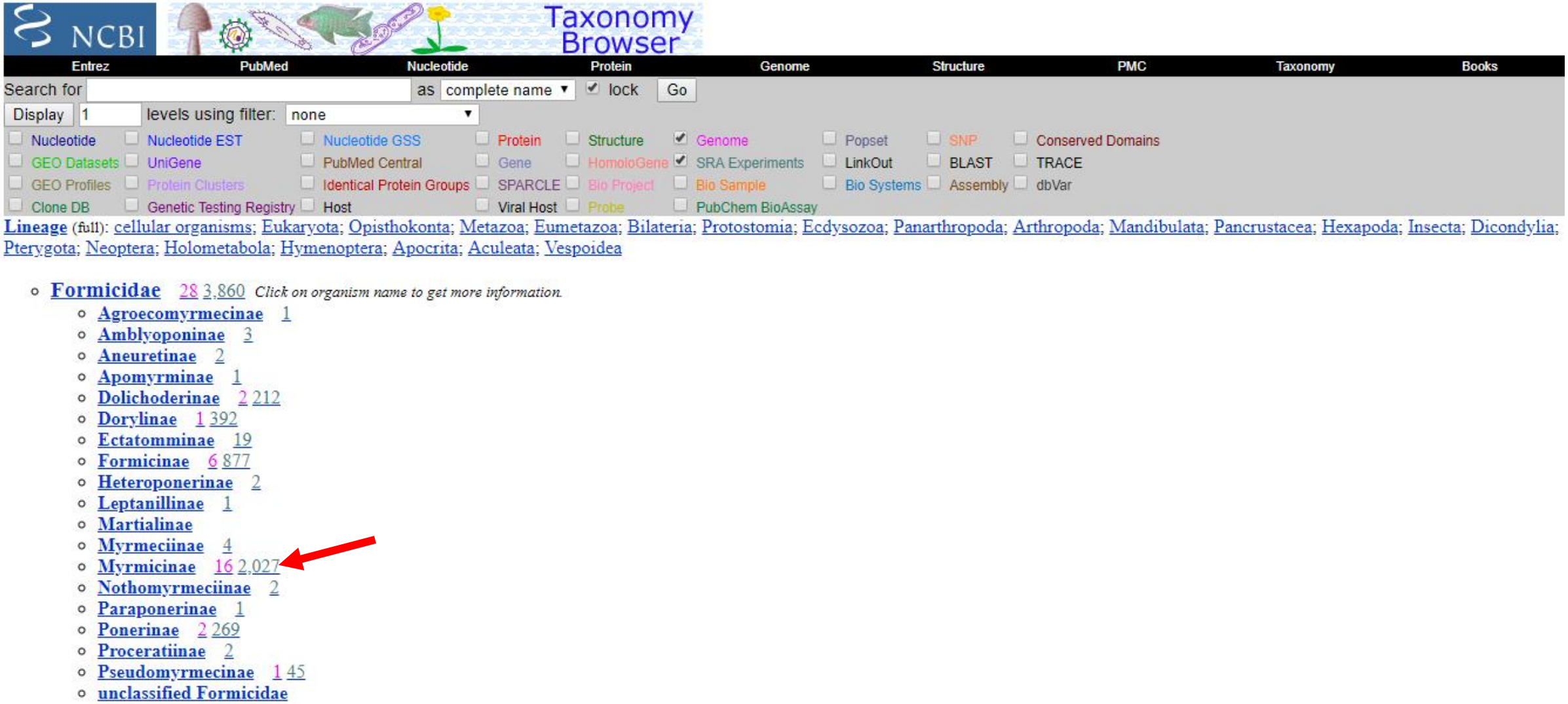
- **Formicidae** *Click on organism name to get more information.*
  - **Agroecomyrmecinae**
    - **Tatuidris**
      - **Tatuidris tatusia**
      - **Tatuidris sp. ECU01**
  - **Amblyoponinae**
    - **Adetomyrma**
      - **Adetomyrma bressleri**
      - **Adetomyrma caputleae**
      - **Adetomyrma sp. 2a CS-2003**
      - **Adetomyrma sp. 2b CSM-2006**
      - **Adetomyrma sp. CS-2003**
      - **Adetomyrma sp. Ma02**
      - **Adetomyrma sp. Ma04**
      - **Adetomyrma sp. MAD02**
    - **Amblyopone**
      - **Amblyopone australis**
      - **Amblyopone longidens**
      - **Amblyopone sp. a398**
      - **Amblyopone sp. CS-2003-1**
      - **Amblyopone sp. CS-2003-2**
      - **Amblyopone sp. CS-2003-6**
      - **Amblyopone sp. CS-2003-7**
      - **Amblyopone sp. CS-2003-8**
      - **Amblyopone sp. CS-2003-9**
      - **Amblyopone sp. Ma01**
    - **Bannapone**
      - **Bannapone scrobiceps**
    - **Concoctio**

- Para facilitar a visualização dos resultados, mude o campo “Display” para 1 e marque os campos “Genome” e “SRA Experiments”. Clique em “Go”

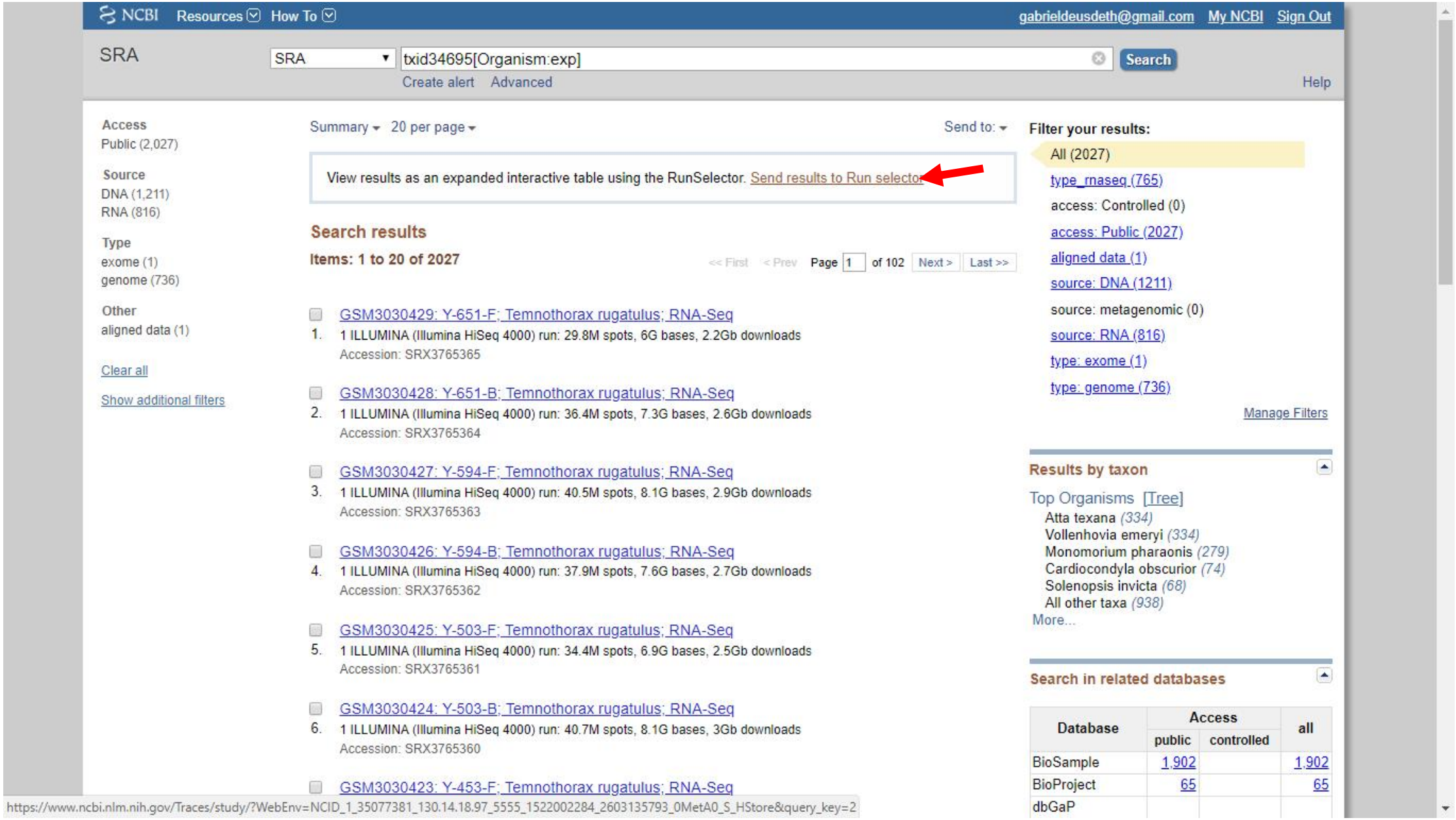
- Agora, além de mais simples, nossa janela de resultados tem 2 números na frente de cada grupo:
  - o rosa diz quantos genomas parciais existem para o grupo
  - o cinza mostra o número de datasets públicos disponíveis



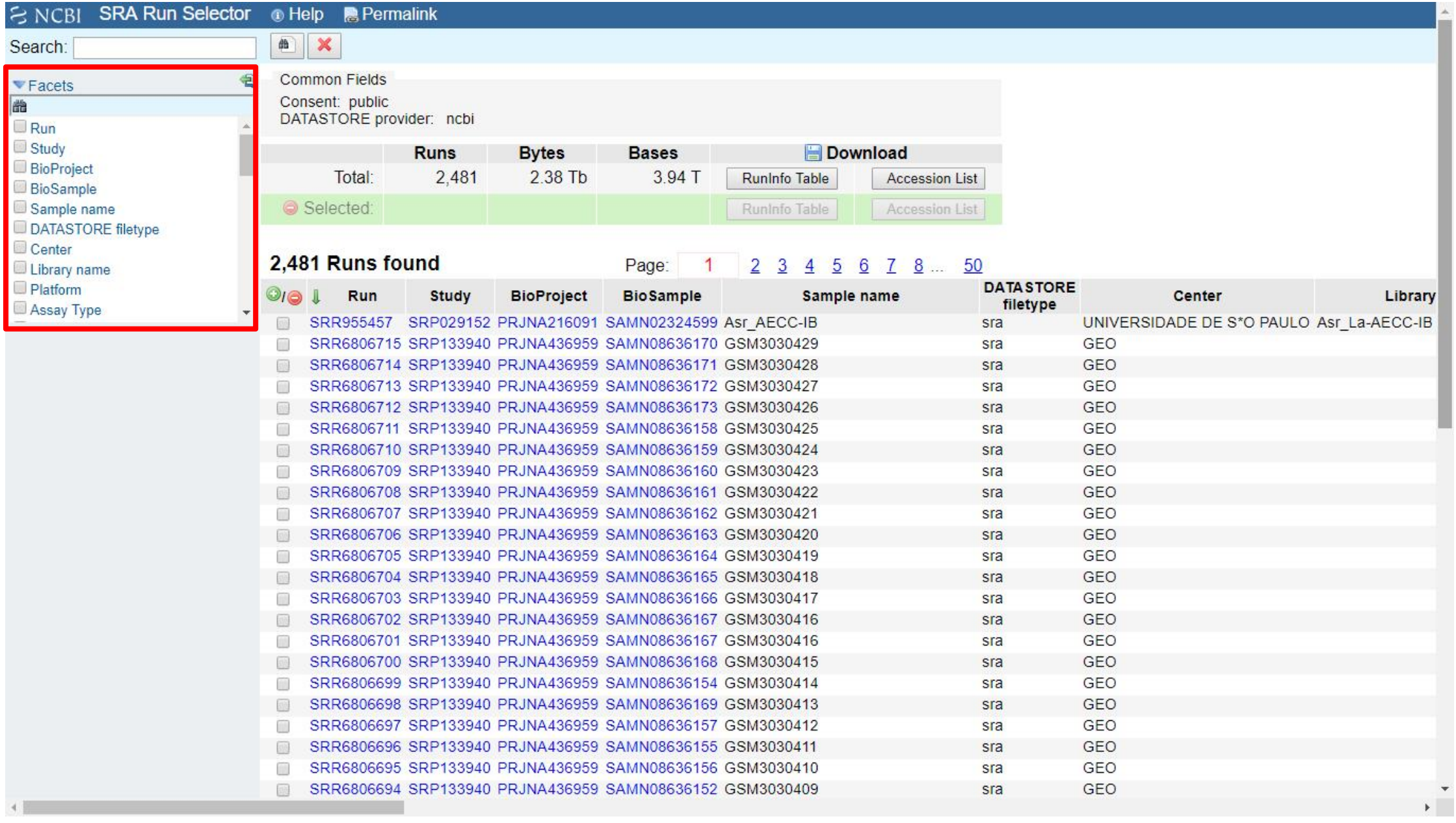
- Como regra geral, um bom grupo para se tentar montar mitocondrias é aquele que possui poucas montagens parciais mas muitos dados públicos
- Nesse exemplo, um bom grupo para se trabalhar é Myrmicinae, que tem apenas 16 genomas parciais, mas mais de 2 mil datasets públicos
- Seguindo o mesmo raciocínio, escolha um grupo e clique no número cinza



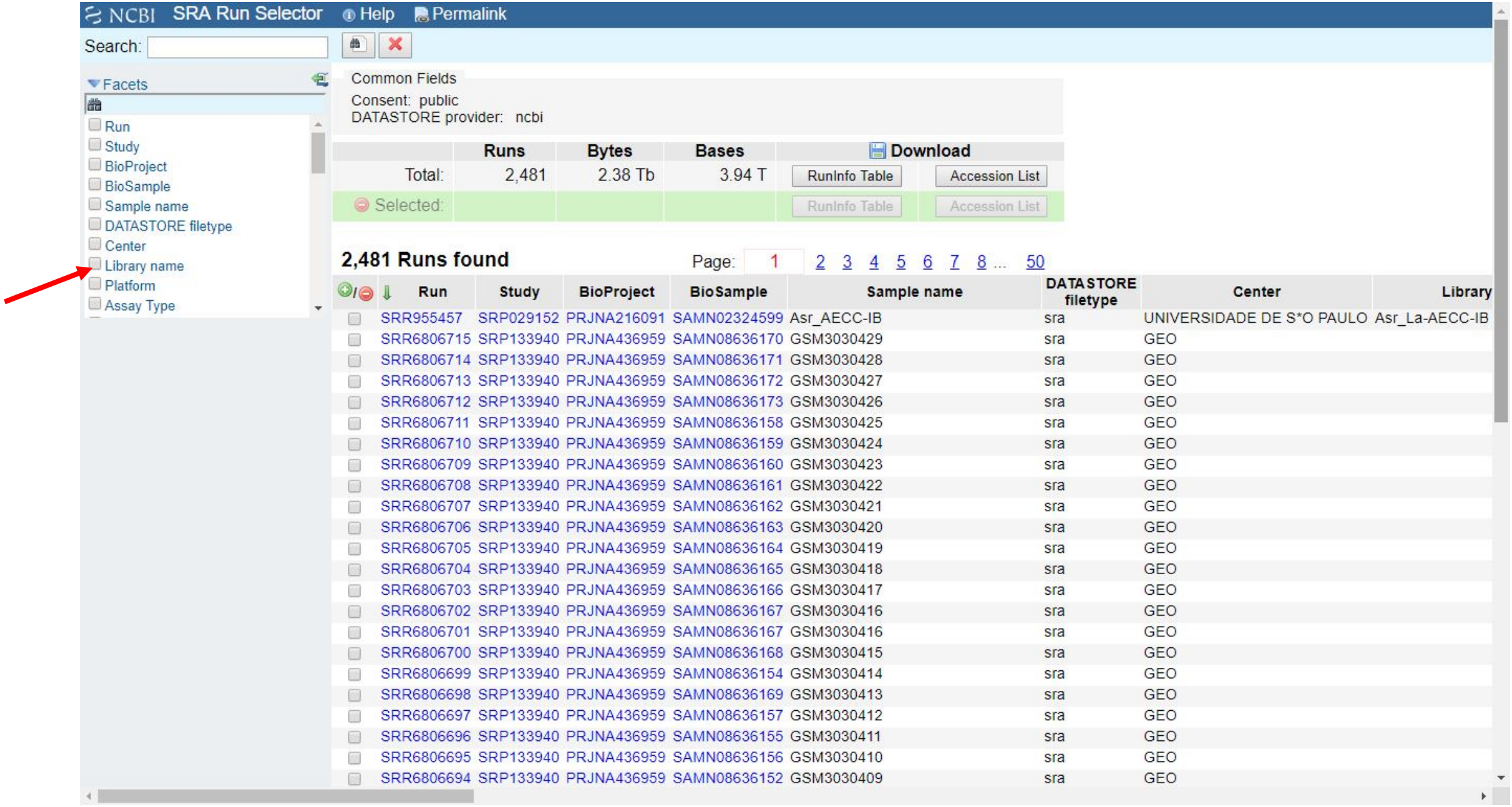
- Na próxima página, clique em “Send results to Run selector”



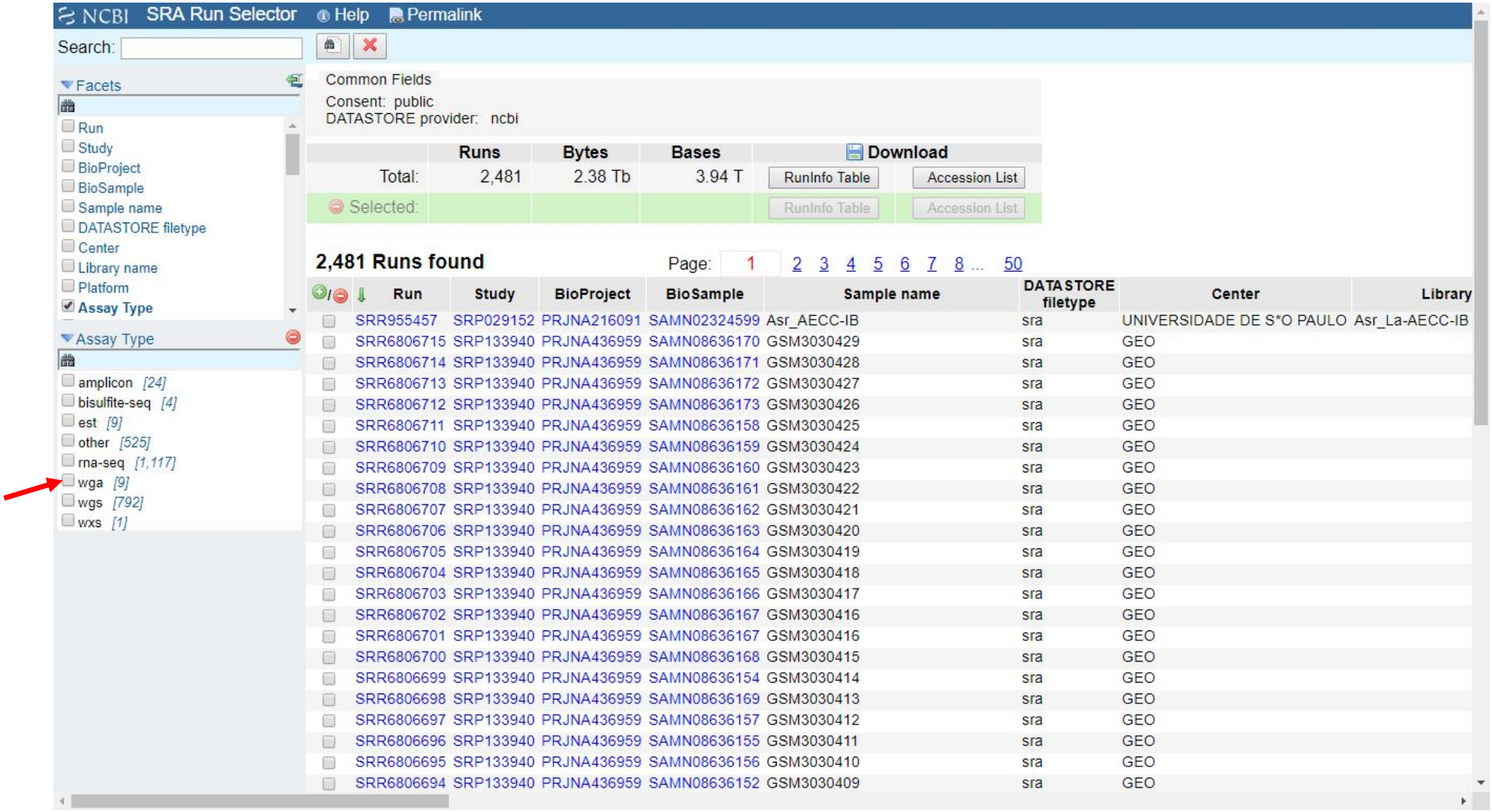
- Agora você está no SRA Run Selector, que te mostra os datasets disponíveis em forma de tabela, onde você pode usar os filtros no canto superior esquerdo da página para navegar melhor por esses dados



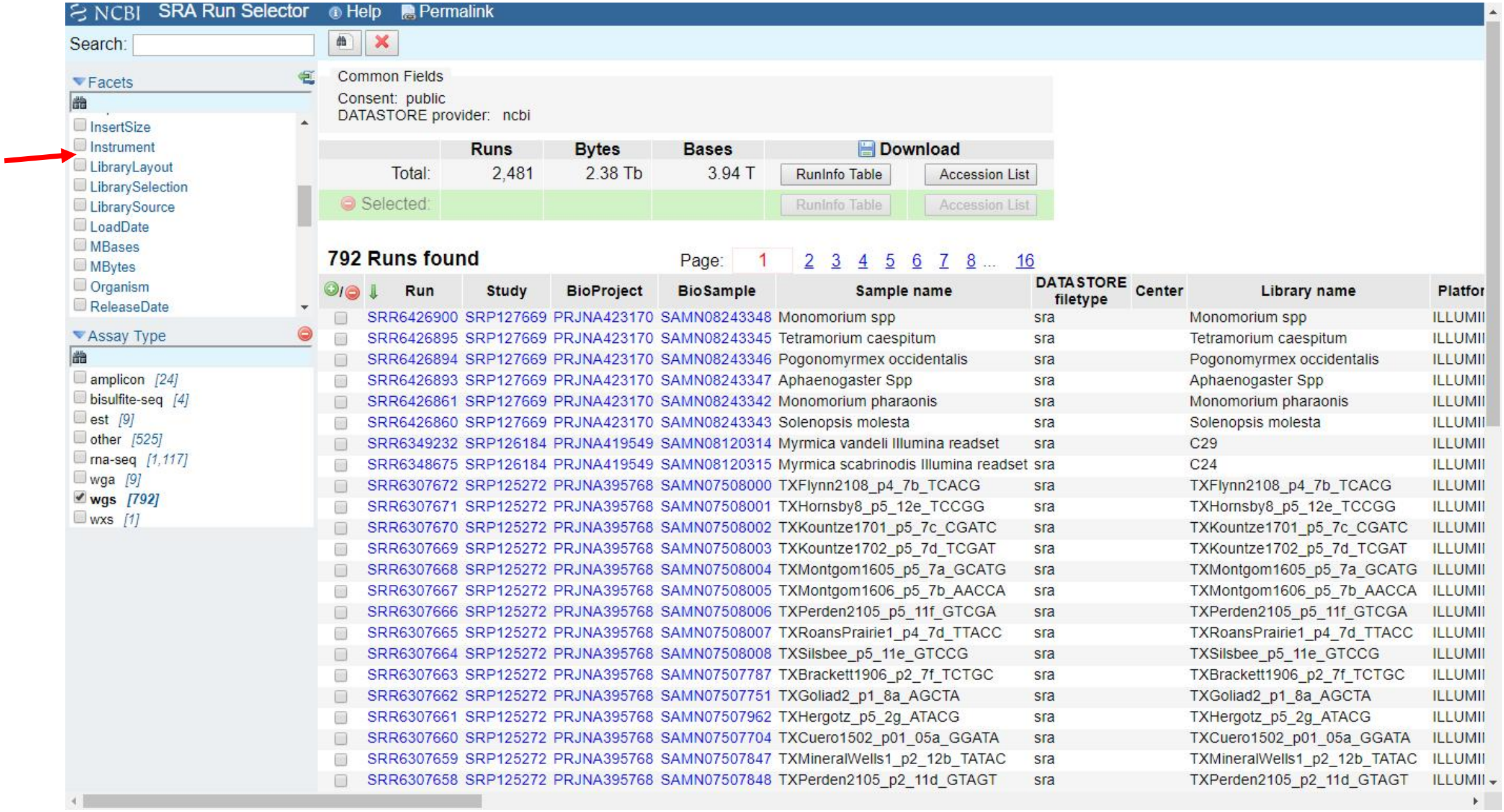
- Nos filtros, clique em “Assay Type”



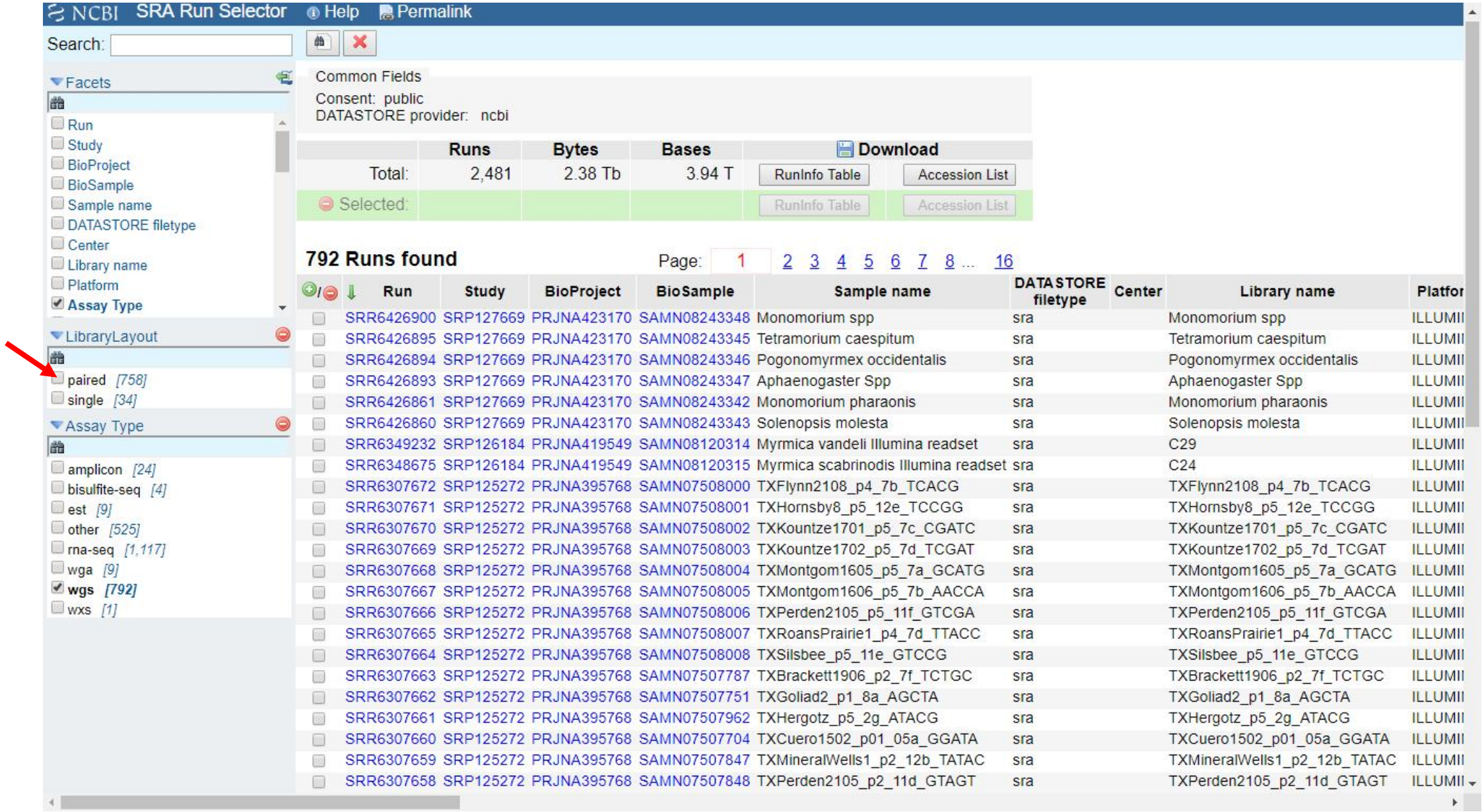
- Clique em “wgs” para selecionar apenas os dados de sequenciamento de DNA total
- (Obs: dados de RNA-seq também podem ser utilizados para montar genomas mitocondriais completos)



- Nos filtros, clique em “Library Layout”



- Selecione “paired”
- (Obs: dados “**single**-end” também podem ser utilizados para obter o genoma mitocondrial completo)



- Agora navegue pela tabela (da esquerda para a direita) até encontrar o campo “Mbytes”
- Clique nele

Instrument	LibraryLayout	Library Selection	Library Source	LoadDate	MBases	MBytes	Organism	ReleaseDate	bioproject id	collection date	cultivar	ecotype
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	4,703	2,171	Monomorium emarginatum	2017-12-28		03-Jul-2008		
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	3,986	1,827	Tetramorium caespitum	2017-12-28		06-Feb-2009		
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	1,141	522	Pogonomyrmex occidentalis	2017-12-28		31-Aug-2008		
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	747	344	Aphaenogaster carolinensis	2017-12-28		31-Aug-2008		
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	632	290	Monomorium pharaonis	2017-12-28		20-Jul-2010		
NextSeq 550	SINGLE	RANDOM	METAGENOMIC	2017-12-28	6,939	3,182	Solenopsis molesta	2017-12-28		22-Apr-2009		
Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,775	7,773	Myrmica vandeli	2017-12-05		AUG-2016		
Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,595	7,620	Myrmica scabrinodis	2017-12-05		AUG-2016		
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	159	78	Atta texana	2017-11-20		Apr-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	549	270	Atta texana	2017-11-20		May-2006	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	355	174	Atta texana	2017-11-20		Mar-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	249	122	Atta texana	2017-11-20		Mar-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	185	91	Atta texana	2017-11-20		Mar-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	311	153	Atta texana	2017-11-20		Mar-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	460	225	Atta texana	2017-11-20		May-2006	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	207	102	Atta texana	2017-11-20		Apr-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	528	259	Atta texana	2017-11-20		Apr-2006	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	603	299	Atta texana	2017-11-20		Mar-2006	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	591	291	Atta texana	2017-11-20		Feb-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	589	291	Atta texana	2017-11-20		Apr-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	431	215	Atta texana	2017-11-20		Mar-2007	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	566	278	Atta texana	2017-11-20		May-2006	not applicable	not applicabl
Illumina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	532	261	Atta texana	2017-11-20		May-2006	not applicable	not applicabl

- Agora, os datasets estão organizados por tamanho, do menor para o maior.
- Clique novamente em “Mbytes”

Instrument	LibraryLayout	LibrarySelection	LibrarySource	LoadDate	MBases	MBytes	Organism	ReleaseDate	bioproject id	collection date	cultivar	ecotype
ina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	6	3	Atta texana	2017-11-20		May-2006	not applicable	not applicable
ina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	13	7	Atta texana	2017-11-20		May-2007	not applicable	not applicable
ina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	14	8	Atta texana	2017-11-20		May-2007	not applicable	not applicable
ina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	23	12	Atta texana	2017-11-20		Apr-2007	not applicable	not applicable
ina HiSeq 4000	PAIRED	Reduced Representation	GENOMIC	2017-11-20	22	12	Atta texana	2017-11-20		Feb-2006	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	77	31	Mycetosoritis hartmanni	2017-04-12			not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2016-11-22	74	32	Sericomyrmex opacus	2017-02-01		25-Sep-2001		
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	88	40	Cyphomyrmex sp. 1 MB-2017	2017-04-14		1992	not applicable	not applicable
ina MiSeq	PAIRED	Hybrid Selection	GENOMIC	2014-05-30	79	41	Sericomyrmex harekulli	2014-05-30		2011		
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	102	44	Lachnomyrmex scrobiculatus	2017-04-12		2008	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	104	45	Mycetagroicus triangularis	2017-04-12		1992	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	113	49	Cyatta abscondita	2017-04-12		2009	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2016-11-22	126	52	Sericomyrmex parvulus	2017-02-01		30-Sep-2008		
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-01-06	107	57	Aphaenogaster phalangium	2017-03-01		2010	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	158	67	Tranopelta gilva	2017-04-12		2006	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2016-11-22	170	69	Sericomyrmex bondari	2017-02-01		19-Nov-2005		
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2017-04-12	114	69	Apterostigma dorotheae	2017-04-12		1996	not applicable	not applicable
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2016-11-22	169	70	Sericomyrmex opacus	2017-02-01		14-Jun-1995		
ina HiSeq 2500	PAIRED	Hybrid Selection	GENOMIC	2016-11-22	116	75	Sericomyrmex mayri	2017-02-01		08-Jan-1995		

- Por fim, os datasets estão organizados do maior para o menor
- (Obs: Geralmente, quanto mais dados disponpiveis, maior a probabilidade de conseguir obter o mitogenoma completo sem problemas. Por isso organizar os dados dessa forma é conveniente ao escolher os dados com os quais se vai trabalhar)

InsertSize	Instrument	LibraryLayout	LibrarySelection	LibrarySource	LoadDate	MBases	MBytes	Organism	ReleaseDate	bioproject id	collection date	cultivar	ecotype
200	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-18	39,051	25,624	Cardiocondyla obscurior	2014-12-01				
200	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-19	37,311	24,281	Cardiocondyla obscurior	2014-12-01				
350	Illumina HiSeq 1000	PAIRED	RANDOM	GENOMIC	2015-11-28	32,731	20,969	Solenopsis invicta	2015-07-22		01-Jan-2010		
500	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-18	29,049	18,789	Cardiocondyla obscurior	2014-12-01				
0	454 GS FLX Titanium	SINGLE	RANDOM	GENOMIC	2012-11-26	4,530	10,445	Solenopsis invicta	2013-02-07		01-Jan-2010		
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-28	14,191	9,874	Trachymyrmex septentrionalis	2016-09-13		2011		
800	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	12,542	8,962	Acromyrmex echinatior	2011-04-27				
5000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	12,147	8,811	Acromyrmex echinatior	2011-04-27				
450	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-10-27	13,657	8,714	Monomorium pharaonis	2014-10-15		missing		
2000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	11,825	8,514	Acromyrmex echinatior	2011-04-27				
10000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	11,165	8,231	Acromyrmex echinatior	2011-04-27				
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-29	11,167	8,016	Cyphomyrmex costatus	2016-09-13		2011		
0	Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,775	7,773	Myrmica vandeli	2017-12-05		AUG-2016		
0	Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,595	7,620	Myrmica scabrinodis	2017-12-05		AUG-2016		
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-07-15	10,185	6,971	Acromyrmex echinatior	2011-04-27				
800	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-29	8,795	6,304	Atta colombica	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-03-03	9,767	5,987	Atta colombica	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-27	8,605	5,627	Cyphomyrmex costatus	2016-09-13		2011		
10000	454 GS FLX Titanium	PAIRED	RANDOM	GENOMIC	2012-11-26	2,040	5,202	Solenopsis invicta	2013-02-07		01-Jan-2010		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-03-03	7,392	4,773	Trachymyrmex septentrionalis	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-28	7,880	4,561	Trachymyrmex cornetzi	2016-09-13		2011		

- Por fim, os datasets estão organizados do maior para o menor
- (Obs: Geralmente, quanto mais dados disponpiveis, maior a probabilidade de conseguir obter o mitogenoma completo sem problemas. Por isso organizar os dados dessa forma é conveniente ao escolher os dados com os quais se vai trabalhar)

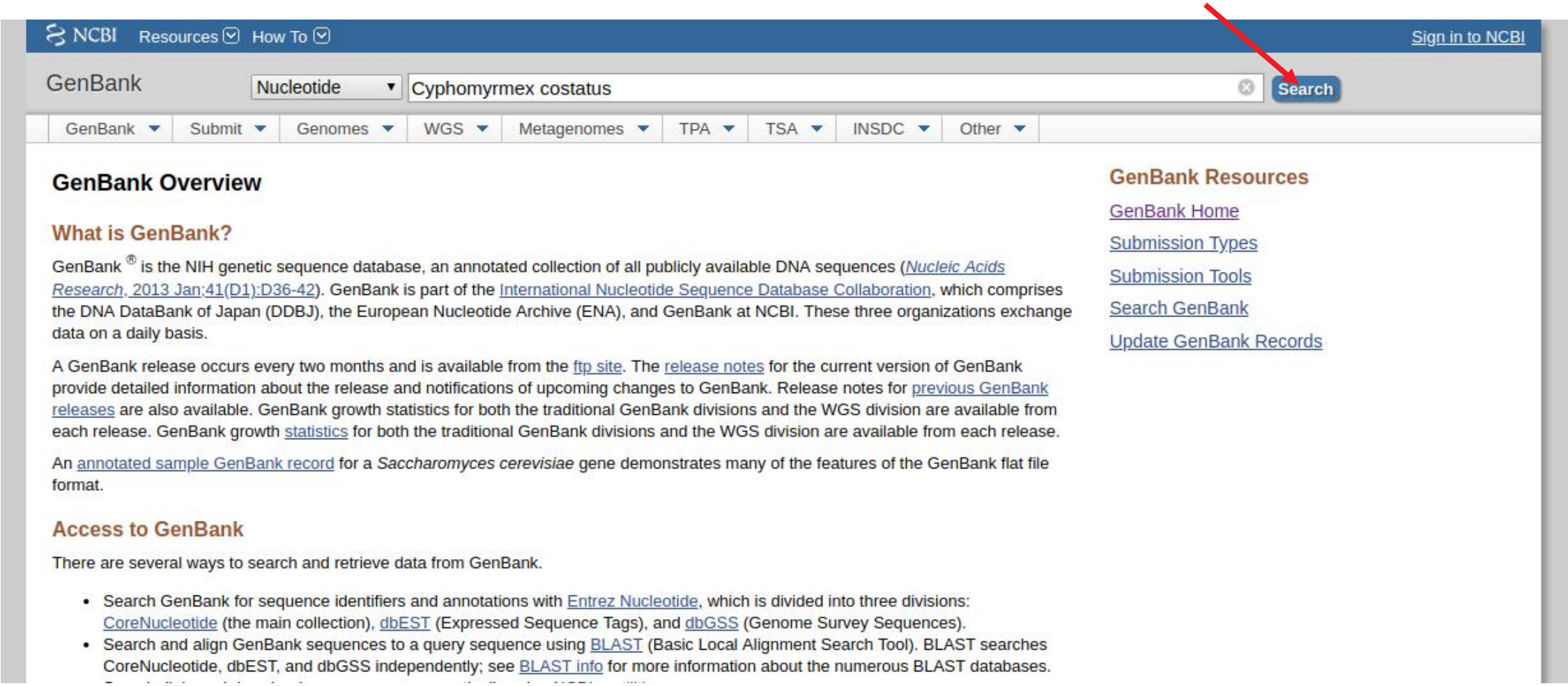
InsertSize	Instrument	LibraryLayout	LibrarySelection	LibrarySource	LoadDate	MBases	MBytes	Organism	ReleaseDate	bioproject id	collection date	cultivar	ecotype
200	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-18	39,051	25,624	Cardiocondyla obscurior	2014-12-01				
200	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-19	37,311	24,281	Cardiocondyla obscurior	2014-12-01				
350	Illumina HiSeq 1000	PAIRED	RANDOM	GENOMIC	2015-11-28	32,731	20,969	Solenopsis invicta	2015-07-22		01-Jan-2010		
500	Illumina HiSeq 2000	PAIRED	unspecified	GENOMIC	2014-09-18	29,049	18,789	Cardiocondyla obscurior	2014-12-01				
0	454 GS FLX Titanium	SINGLE	RANDOM	GENOMIC	2012-11-26	4,530	10,445	Solenopsis invicta	2013-02-07		01-Jan-2010		
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-28	14,191	9,874	Trachymyrmex septentrionalis	2016-09-13		2011		
800	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	12,542	8,962	Acromyrmex echinatior	2011-04-27				
5000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	12,147	8,811	Acromyrmex echinatior	2011-04-27				
450	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-10-27	13,657	8,714	Monomorium pharaonis	2014-10-15		missing		
2000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	11,825	8,514	Acromyrmex echinatior	2011-04-27				
10000	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-05-30	11,165	8,231	Acromyrmex echinatior	2011-04-27				
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-29	11,167	8,016	Cyphomyrmex costatus	2016-09-13		2011		
0	Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,775	7,773	Myrmica vandeli	2017-12-05		AUG-2016		
0	Illumina HiSeq 2500	PAIRED	PCR	GENOMIC	2017-12-05	20,595	7,620	Myrmica scabrinodis	2017-12-05		AUG-2016		
500	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2014-07-15	10,185	6,971	Acromyrmex echinatior	2011-04-27				
800	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-29	8,795	6,304	Atta colombica	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-03-03	9,767	5,987	Atta colombica	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-27	8,605	5,627	Cyphomyrmex costatus	2016-09-13		2011		
10000	454 GS FLX Titanium	PAIRED	RANDOM	GENOMIC	2012-11-26	2,040	5,202	Solenopsis invicta	2013-02-07		01-Jan-2010		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-03-03	7,392	4,773	Trachymyrmex septentrionalis	2016-09-13		2011		
200	Illumina HiSeq 2000	PAIRED	RANDOM	GENOMIC	2016-02-28	7,880	4,561	Trachymyrmex cornetzi	2016-09-13		2011		

- Outras dicas para selcionar seu dataset:
  - **AvgSpotLen**: É o dobro do tamanho das reads. Quanto maior, melhor.
  - **LibrarySelection**: RANDOM, esperam-se reads mitocondriais em maior quantidade
  - **LibrarySource**: GENOMIC, apesat de também ser possível utilizar dados metagenômicos
  - **ReleaseDate**: Quando o dataset se tornou público. Evitar datasets muito recentes, já que talvez aqueles que produziram as sequências tenham interesse em fechar o DNA mitocondrial

- Assim que escolher seu dataset, veja a que organismo ele pertence no campo Organism

LibraryLayout	LibrarySelection	LibrarySource	LoadDate	MBases	MBytes	Organism	ReleaseDate	bioproject id	collection date	cultivar	ecotype	env biome	env feature	env material	estima size
PAIRED	unspecified	GENOMIC	2014-09-18	39,051	25,624	Cardiocondyla obscurior	2014-12-01								
PAIRED	unspecified	GENOMIC	2014-09-19	37,311	24,281	Cardiocondyla obscurior	2014-12-01								
PAIRED	RANDOM	GENOMIC	2015-11-28	32,731	20,969	Solenopsis invicta	2015-07-22		01-Jan-2010						
PAIRED	unspecified	GENOMIC	2014-09-18	29,049	18,789	Cardiocondyla obscurior	2014-12-01								
SINGLE	RANDOM	GENOMIC	2012-11-26	4,530	10,445	Solenopsis invicta	2013-02-07		01-Jan-2010						
PAIRED	RANDOM	GENOMIC	2016-02-28	14,191	9,874	Trachymyrmex septentrionalis	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2014-05-30	12,542	8,962	Acromyrmex echinatior	2011-04-27								
PAIRED	RANDOM	GENOMIC	2014-05-30	12,147	8,811	Acromyrmex echinatior	2011-04-27								
PAIRED	RANDOM	GENOMIC	2014-10-27	13,657	8,714	Monomorium pharaonis	2014-10-15		missing			missing	missing	missing	missing
PAIRED	RANDOM	GENOMIC	2014-05-30	11,825	8,514	Acromyrmex echinatior	2011-04-27								
PAIRED	RANDOM	GENOMIC	2014-05-30	11,165	8,231	Acromyrmex echinatior	2011-04-27								
PAIRED	RANDOM	GENOMIC	2016-02-29	11,167	8,016	Cyphomyrmex costatus	2016-09-13		2011						
PAIRED	PCR	GENOMIC	2017-12-05	20,775	7,773	Myrmica vandeli	2017-12-05		AUG-2016						
PAIRED	PCR	GENOMIC	2017-12-05	20,595	7,620	Myrmica scabrinodis	2017-12-05		AUG-2016						
PAIRED	RANDOM	GENOMIC	2014-07-15	10,185	6,971	Acromyrmex echinatior	2011-04-27								
PAIRED	RANDOM	GENOMIC	2016-02-29	8,795	6,304	Atta colombica	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-03-03	9,767	5,987	Atta colombica	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-02-27	8,605	5,627	Cyphomyrmex costatus	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2012-11-26	2,040	5,202	Solenopsis invicta	2013-02-07		01-Jan-2010						
PAIRED	RANDOM	GENOMIC	2016-03-03	7,392	4,773	Trachymyrmex septentrionalis	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-02-28	7,880	4,561	Trachymyrmex cornetzi	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-03-03	8,379	4,558	Trachymyrmex cornetzi	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-02-26	7,702	4,502	Trachymyrmex cornetzi	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2016-02-25	8,024	4,312	Trachymyrmex zeteki	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2012-11-28	5,808	3,894	Solenopsis invicta	2013-02-07								
PAIRED	RANDOM	GENOMIC	2016-02-25	6,446	3,676	Trachymyrmex zeteki	2016-09-13		2011						
PAIRED	RANDOM	GENOMIC	2012-11-28	5,452	3,613	Solenopsis invicta	2013-02-07								
PAIRED	RANDOM	GENOMIC	2012-11-28	5,388	3,541	Solenopsis invicta	2013-02-07								
PAIRED	RANDOM	GENOMIC	2012-11-28	5,356	3,526	Solenopsis invicta	2013-02-07								
PAIRED	RANDOM	GENOMIC	2016-03-01	5,189	3,407	Atta colombica	2016-09-13		2011						

Agora vamos checar se há mitogenoma completo para nosso Pokémon. Em outra aba, abra o Genbank (https://www.ncbi.nlm.nih.gov/genbank/) e digite o nome do organismo selecionado. Clique em “Search”.



Todas as sequencias disponíveis para o organismo vão aparecer. Entretanto, queremos selecionar apenas as mitocondriais. Para isso, precisamos clicar em “Show additional filters”...

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide Cyphomyrmex costatus Search

Create alert Advanced Help

Species
Animals (33,560)
Fungi (58)
Bacteria (8)
Customize ...

Molecule types
genomic DNA/RNA (17,544)
mRNA (15,690)
Customize ...

Source databases
INSDC (GenBank) (2,165)
RefSeq (31,461)
Customize ...

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

Summary 20 per page Sort by Default order Send to:

Items: 1 to 20 of 33626

<< First < Prev Page 1 of 1682 Next > Last >>

1. PREDICTED: **Cyphomyrmex costatus** pancreas transcription factor 1 subunit alpha-like (LOC108782334), mRNA
786 bp linear mRNA
Accession: XM_018550592.1 GI: 1070619054
BioProject Protein Taxonomy
GenBank FASTA Graphics

2. PREDICTED: **Cyphomyrmex costatus** DNA excision repair protein haywire-like (LOC108782333), mRNA
2,861 bp linear mRNA
Accession: XM_018550591.1 GI: 1070619052
BioProject Protein Taxonomy
GenBank FASTA Graphics

3. PREDICTED: **Cyphomyrmex costatus** serine/threonine-protein kinase B-raf (LOC108782331), transcript variant X2, mRNA
4,384 bp linear mRNA
Accession: XM_018550590.1 GI: 1070619050
BioProject Protein Taxonomy
GenBank FASTA Graphics

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]
Cyphomyrmex costatus *(33560)*
Cyphomyrmex costatus fungal symbiont NJM-2012 *(19)*
Leucoagaricus gongylophorus *(6)*
Leucoagaricus sp. S77 *(3)*
Escovopsis sp. CC1 *(3)*
All other taxa *(35)*
More...

Find related data

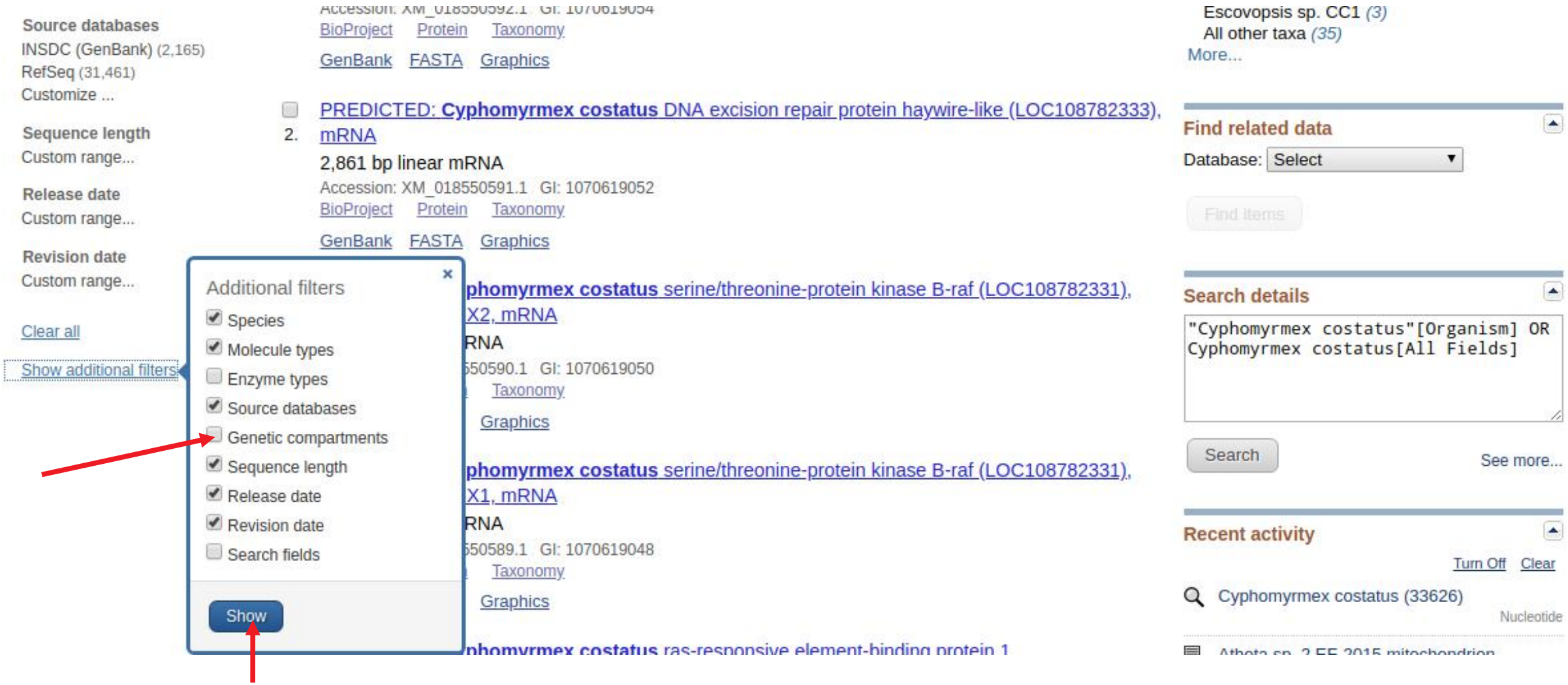
Database: Select

Find items

Search details

```
"Cyphomyrmex costatus"[Organism] OR
Cyphomyrmex costatus[All Fields]
```

...e selecionar “Genetic compartiments”. Logo depois, clique em “Show”.



Novos filtros aparecerão, dentre eles o de sequencias mitocondriais. Clique em “Mitochondrion”.

NCBI Resources How To Sign in to NCBI

Nucleotide | Nucleotide | Cyphomyrmex costatus | Search

Create alert Advanced Help

Species
Animals (33,560)
Fungi (58)
Bacteria (8)
Customize ...

Molecule types
genomic DNA/RNA (17,544)
mRNA (15,690)
Customize ...

Source databases
INSDC (GenBank) (2,165)
RefSeq (31,461)
Customize ...

Genetic compartments
Chloroplast
Mitochondrion
Plasmid
Plastid

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

Show additional filters

Summary 20 per page Sort by Default order Send to:

Items: 1 to 20 of 33626

<< First < Prev Page 1 of 1682 Next > Last >>

1. PREDICTED: **Cyphomyrmex costatus** pancreas transcription factor 1 subunit alpha-like (LOC108782334), mRNA
786 bp linear mRNA
Accession: XM_018550592.1 GI: 1070619054
BioProject Protein Taxonomy
GenBank FASTA Graphics

2. PREDICTED: **Cyphomyrmex costatus** DNA excision repair protein haywire-like (LOC108782333), mRNA
2,861 bp linear mRNA
Accession: XM_018550591.1 GI: 1070619052
BioProject Protein Taxonomy
GenBank FASTA Graphics

3. PREDICTED: **Cyphomyrmex costatus** serine/threonine-protein kinase B-raf (LOC108782331), transcript variant X2, mRNA
4,384 bp linear mRNA
Accession: XM_018550590.1 GI: 1070619050
BioProject Protein Taxonomy
GenBank FASTA Graphics

4. PREDICTED: **Cyphomyrmex costatus** serine/threonine-protein kinase B-raf (LOC108782331), transcript variant X1, mRNA
4,492 bp linear mRNA

Filters: Manage Filters

Results by taxon

Top Organisms [Tree]
Cyphomyrmex costatus *(33560)*
Cyphomyrmex costatus fungal symbiont NJM-2012 *(19)*
Leucoagaricus gongylophorus *(6)*
Leucoagaricus sp. S77 *(3)*
Escovopsis sp. CC1 *(3)*
All other taxa *(35)*
More...

Find related data

Database: Select

Find items

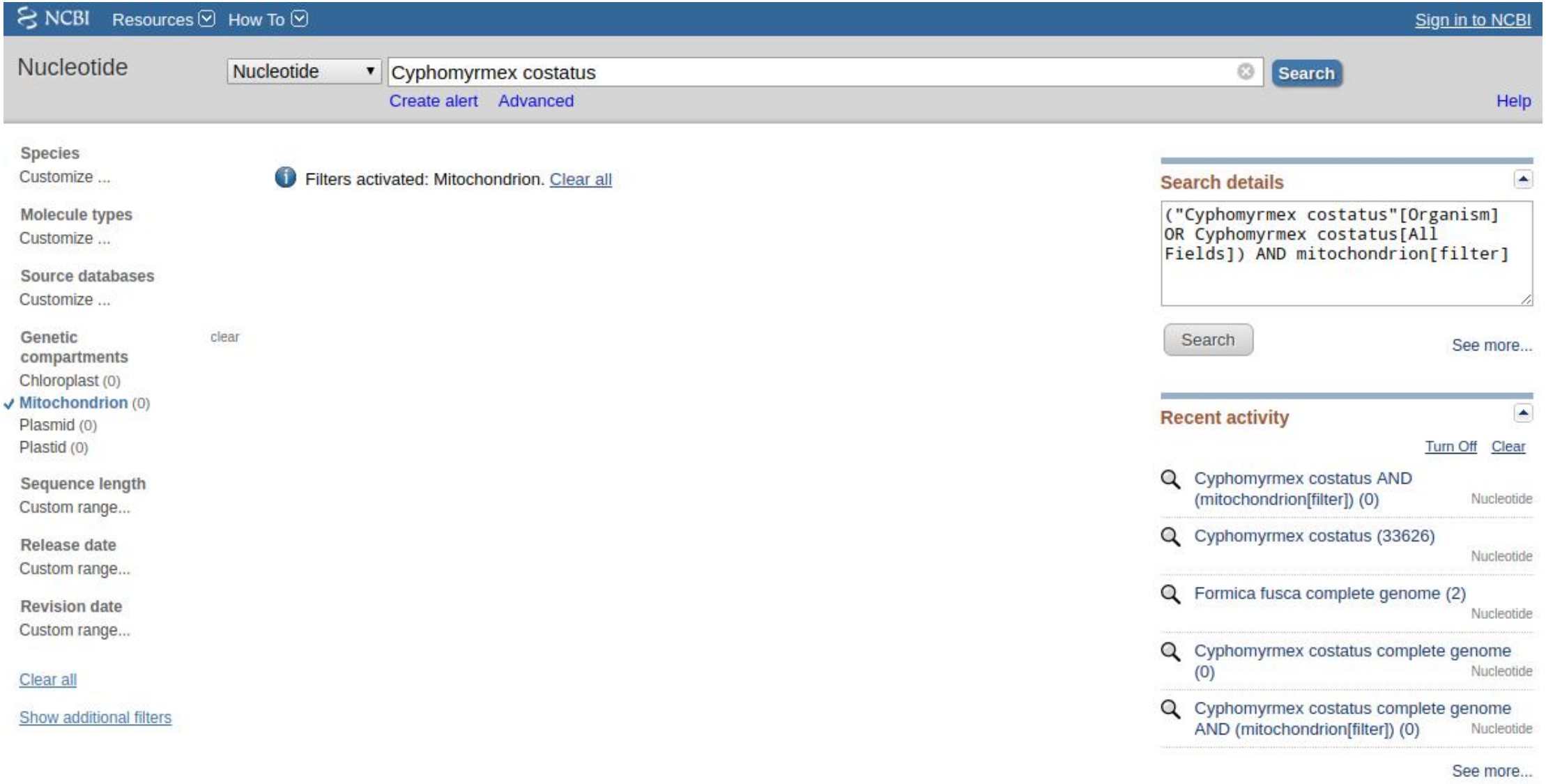
Search details

```
"Cyphomyrmex costatus"[Organism] OR
Cyphomyrmex costatus[All Fields]
```

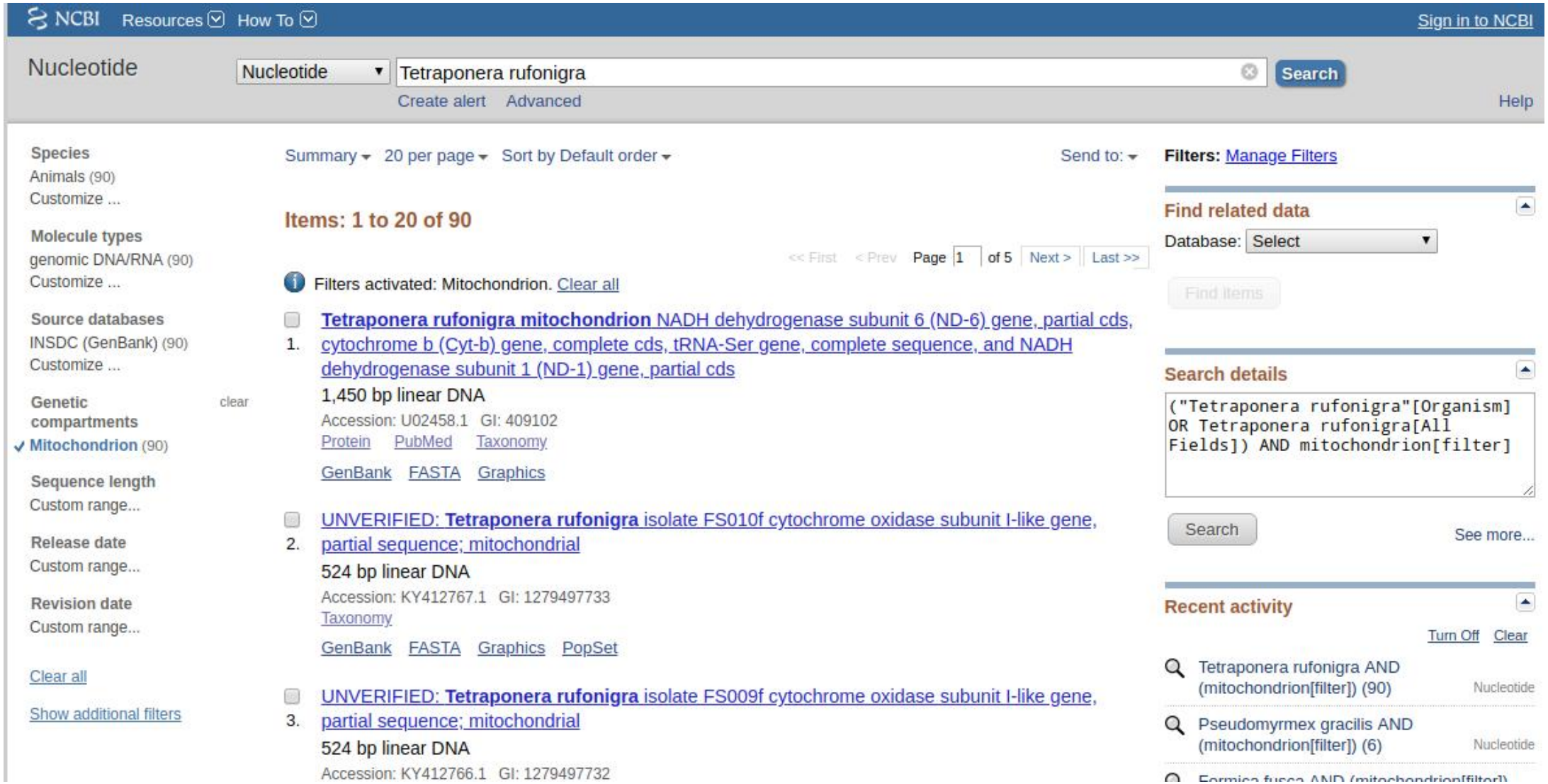
Search See more...

Recent activity

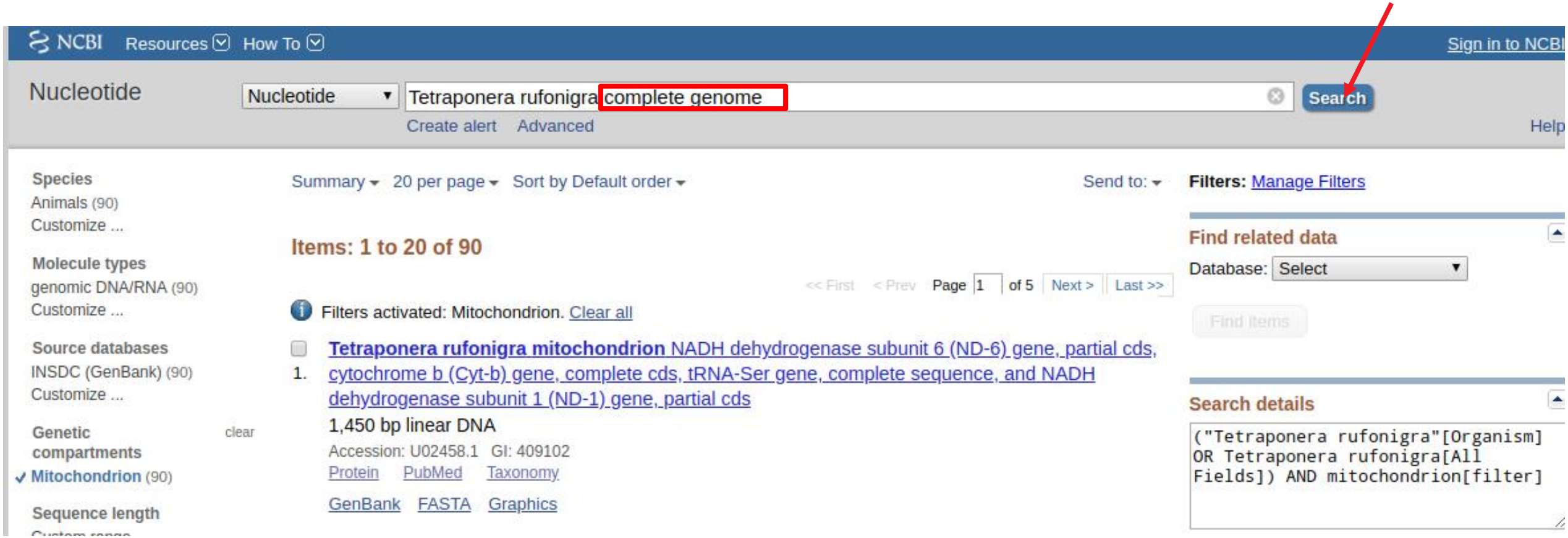
Nesse caso, não há nenhuma sequência mitocondrial montada para essa espécie e, consequentemente, não há mitogenoma completo para ela. Logo, essa é uma boa espécie para se trabalhar, já que geraremos pela primeira vez a sequência mitocondrial completa para ela.



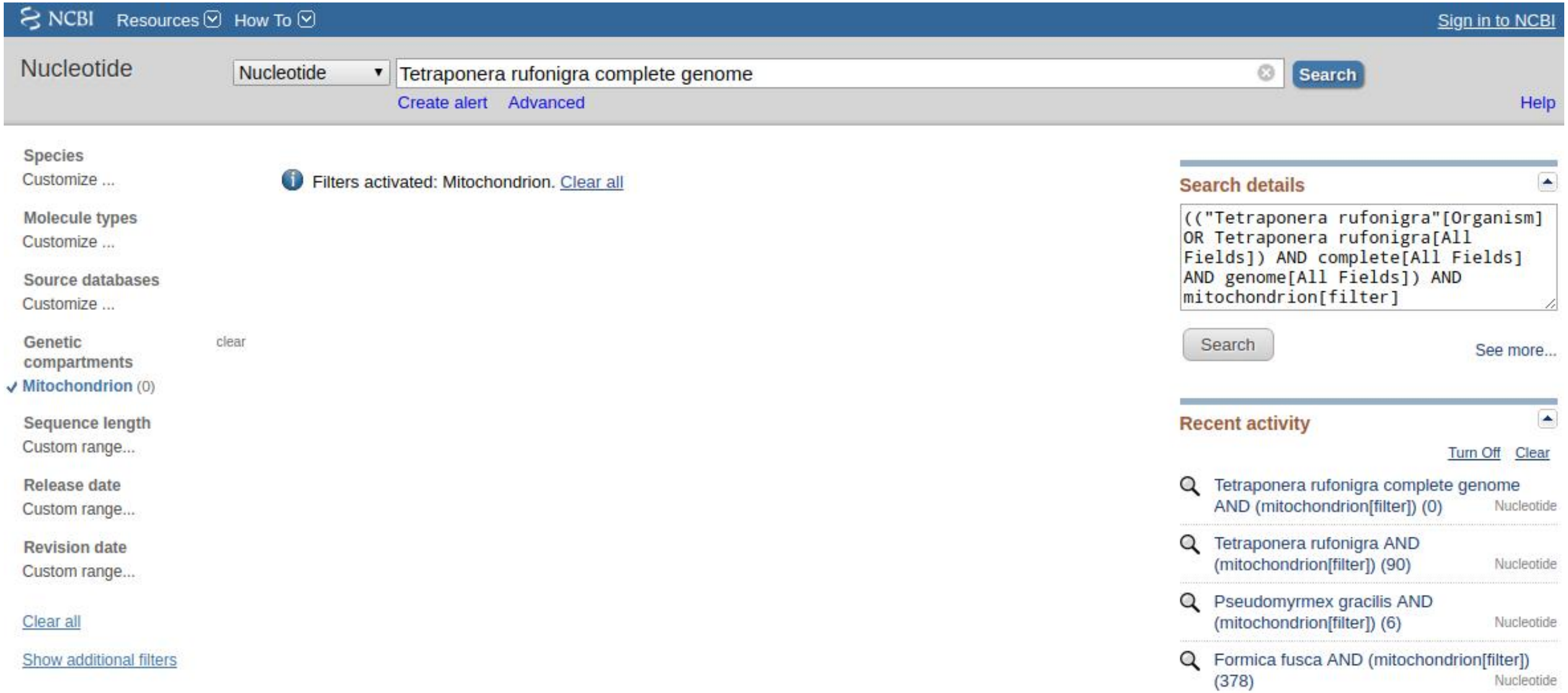
Entretanto, às vezes aparecerão muitos resultados de genes parciais, como é o caso da formiga *Tetraponera rufonigra*.Em vez de checar um por um, podemos escrever “complete genome” na frente da espécie escolhida e procurar por resultados.



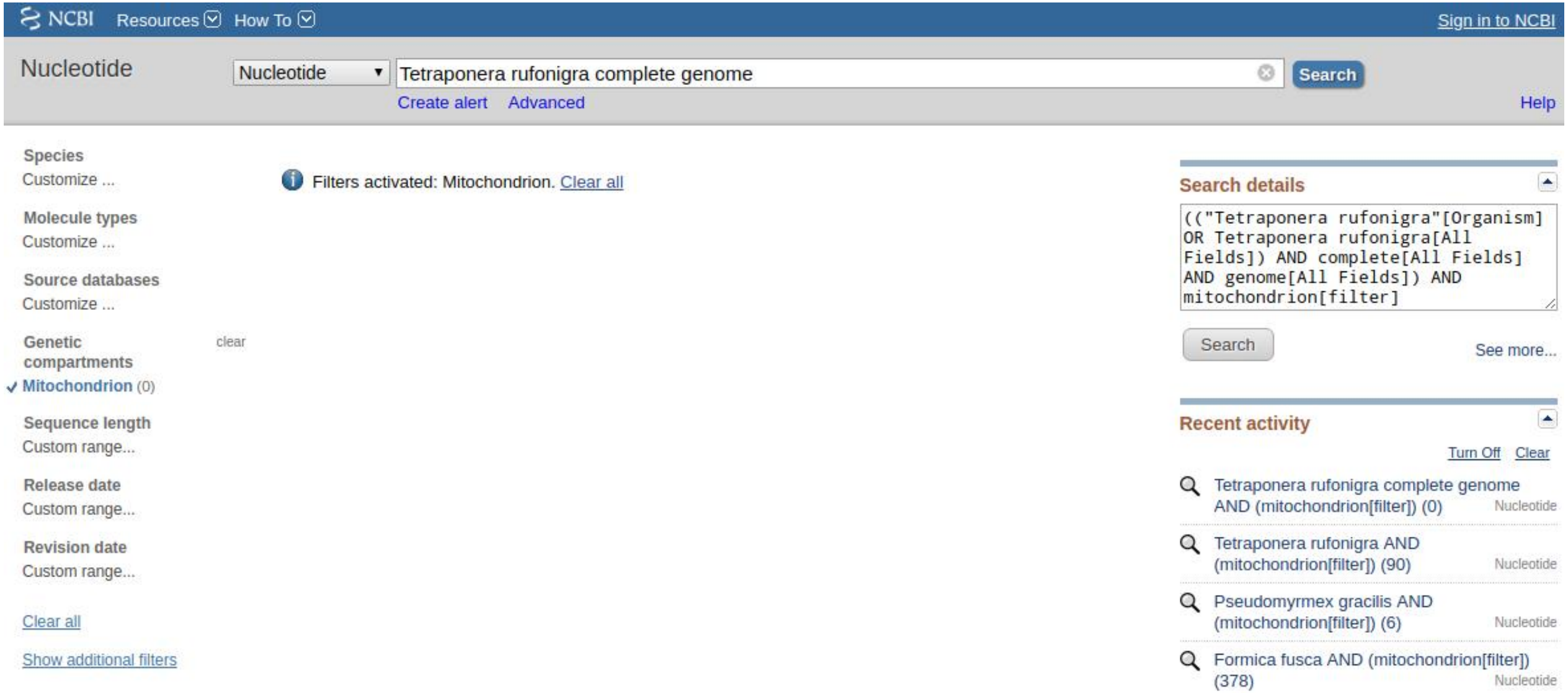
Em vez de checar um por um, podemos escrever “complete genome” na frente do nome da espécie escolhida e clicar em “Search” novamente.



Se nenhum resultado for encontrado, então não há mitogenoma completo descrito para a espécie e nós podemos montá-lo pela primeira vez. Por outro lado, se a mitocondria for encontrada, devemos selecionar outro organismo.



Se nenhum resultado for encontrado, então não há mitogenoma completo descrito para a espécie e nós podemos montá-lo pela primeira vez. Por outro lado, se a mitocondria for encontrada, devemos selecionar outro organismo.



Voltando à aba do SRA Run Selector, vá até a primeira coluna (“Run”, que para nós pode ser considerado um sinônimo de dataset) e copie o código do dataset que você selecionou. No caso, é a “SRR3168941”.

Obs: Esses códigos das runs sempre começam com “SRR”, “ERR” ou “DRR”. Não confundir com os códigos dos campos “Study”, “Bioproject” e “BioSample”.

Facets

- Run
- Study
- BioProject
- BioSample
- Sample name
- DATASTORE filetype
- Center
- Library name
- Platform
- **Assay Type**

Assay Type

- **wgs** *[799]*

LibraryLayout

- **paired** *[799]*
- single *[34]*

Common Fields

Consent: public
DATASTORE provider: ncbi

	Runs	Bytes	Bases	Download	
Total:	2,916	2.54 Tb	4.22 T	RunInfo Table	Accession List
Selected:				RunInfo Table	Accession List

799 Runs found

Page: 1 2 3 4 5 6 7 8 ... 16

Run	Study	BioProject	BioSample	Sample name	DATASTORE filetype	Center
SRR1578062	SRP046084	PRJNA237579	SAMN03020263	JP_oypB	sra	UNIVERSITY OF REGENSBURG
SRR1564444	SRP046084	PRJNA237579	SAMN02687466	BR_alpha09	sra	UNIVERSITY OF REGENSBURG
SRR621118	SRP017299	PRJNA182124	SAMN01818676	Single Gp-9b male - DonatelloZ8	sra	QUEEN MARY UNIVERSITY OF LO
SRR1565416	SRP046084	PRJNA237579	SAMN02687466	BR_alpha09	sra	UNIVERSITY OF REGENSBURG
SRR3168933	SRP070115	PRJNA292625	SAMN03982881	Tsep_0110901987	sra	BGI-SHENZHEN
ERR034187	ERP000666	PRJEB2532	SAMEA762107	SAMEA762107	sra	BGI
ERR034189	ERP000666	PRJEB2532	SAMEA762108	SAMEA762108	sra	BGI
DRR023370	DRP002363	PRJDB3164	SAMD00020277	SAMD00020277	sra	OIST (OKINAWA INSTITUTE OF SC
ERR034188	ERP000666	PRJEB2532	SAMEA762109	SAMEA762109	sra	BGI
ERR034190	ERP000666	PRJEB2532	SAMEA762110	SAMEA762110	sra	BGI
SRR3168941	SRP070116	PRJNA292630	SAMN03982885	Ccos_0120105061	sra	BGI-SHENZHEN
SRR6349232	SRP126184	PRJNA419549	SAMN08120314	Myrmica vandeli Illumina readset	sra	
SRR6348675	SRP126184	PRJNA419549	SAMN08120315	Myrmica scabrinodis Illumina readset	sra	

Agora, abra uma nova aba e entre no European Nucleotide Archive – ENA (https://www.ebi.ac.uk/ena).

EMBL-EBI

Services | Research | Training | About us

ENA
European Nucleotide Archive

Search

Examples: BN000065, histone

Advanced
Sequence

Home | Search & Browse | Submit & Update | Software | About ENA | Support

European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. More about ENA

Access to ENA data is provided though the browser, through search tools, large scale file download and through the API.

Text Search

Examples: BN000065, histone

Search

Advanced search

Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

Popular

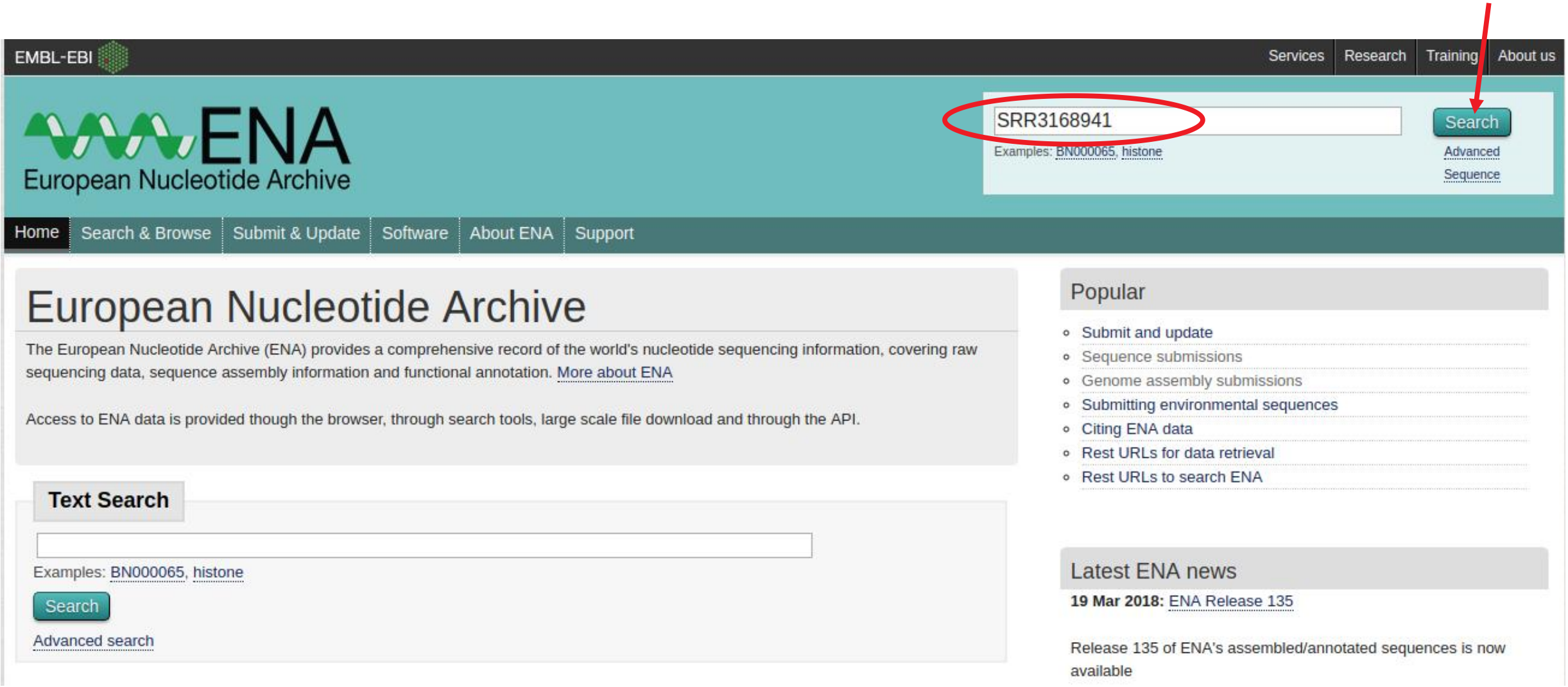
- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- Rest URLs for data retrieval
- Rest URLs to search ENA

Latest ENA news

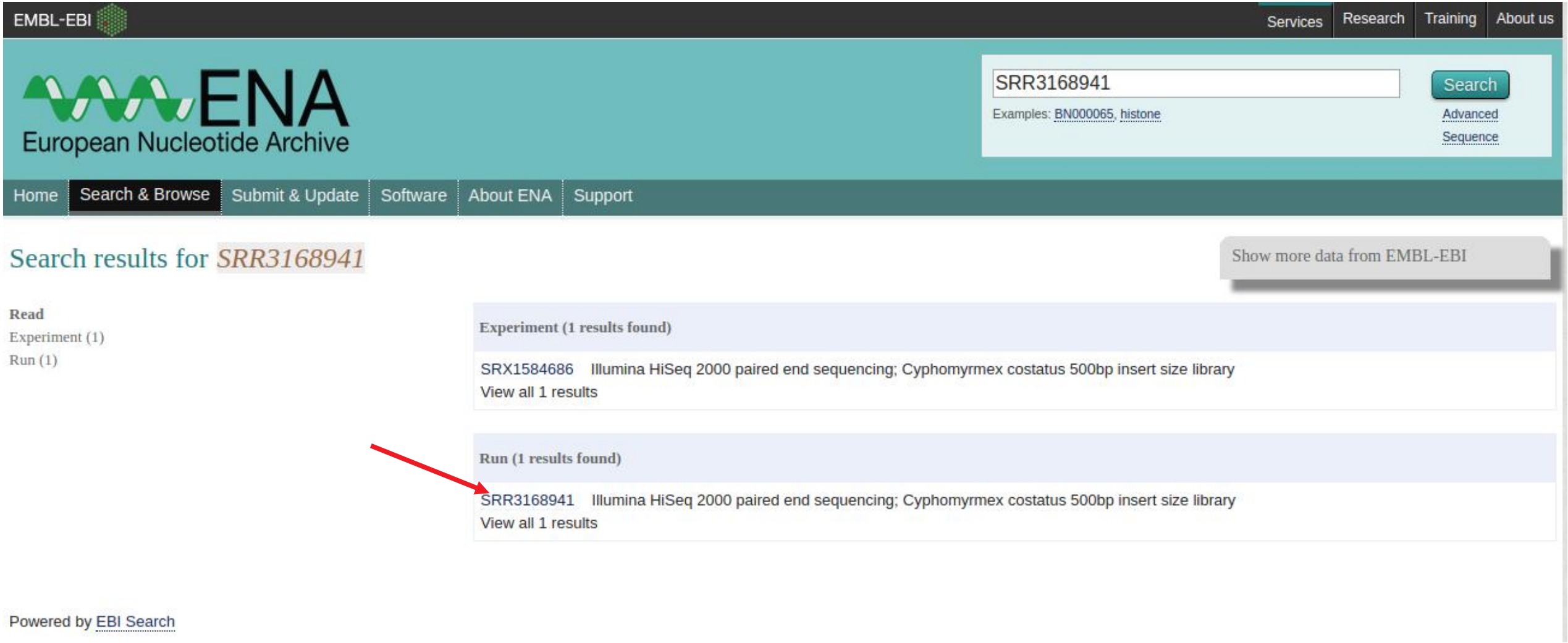
19 Mar 2018: ENA Release 135

Release 135 of ENA's assembled/annotated sequences is now available

No canto superior direito, cole o número do dataset e clique em search.



Clique no número da Run.



Copie o link que está abaixo de “NCBI SRA file (FTP)”

Contact Helpdesk

Run: SRR3168941

Illumina HiSeq 2000 paired end sequencing; Cyphomyrmex costatus 500bp insert size library

View: XML

Download: XML

Submitting Centre	Platform	Model	Read Count	Base Count
	ILLUMINA	Illumina HiSeq 2000	58,548,985	11,709,797,000

Library Layout	Library Strategy	Library Source	Library Selection	Library Name
PAIRED	WGS	GENOMIC	RANDOM	

Broker Name
NCBI

Navigation | Read Files

This table contains the files for run SRR3168941

Bulk Download Files (Using Firefox is recommended to launch the bulk downloader app.)

Download: 1 - 1 of 1 results in TEXT

Select columns

Showing results 1 - 1 of 1 results

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRAM Index files (FTP)	CRAM Index files (Galaxy)
PRJNA292630	SAMN03982885	SRS1296252	SRX1584686	SRR3168941	456900	Cyphomyrmex costatus	Illumina HiSeq 2000	PAIRED	File 1 File 2	File 1 File 2			File 1	File 1		

Hora de entrar na Matrix! Abra o seu terminal (Ctrl+Alt+T).

Caso você não esteja realizando realizando os procedimentos localmente, mas em um servidor:

Use o comando “ssh” para acessar o servidor.

Digite: ssh seu_nome_de_usuário@ip_do_servidor

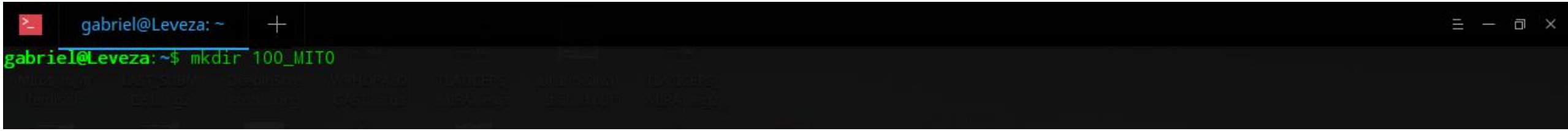
E aperte “Enter”

Obs: o ssh é case-sensitive, diferenciando minúsculas de maiúsculas. Ou seja, se meu nome de usuário é “**g**abriel”, e eu usar “**G**abriel” no ssh, vai dar errado por causa da maiúscula.

Em seguida, digite sua senha e vc entrará no servidor

Recomendo criar uma pasta no seu home para baixar os dados e realizar as montagens. Para fazer isso, basta usar o comando “mkdir” (make directory)

Digite: mkdir nome_da_pasta



Agora entre na pasta que criou usando o comando “cd” (change directory)

Digite: cd nome_da_pasta

Agora, podemos baixar o dataset selecionado dentro dessa pasta. Para isso, vamos usar o comando “wget”. Entretanto, vamos também mudar o nome do arquivo baixado (usando a opção “-O” (é um “Ó” maiúsculo, não um zero)) e redirecionar saída (usando o “>”) e erro (usando “2>”) padrão para depois termos certeza que o download foi completado.

Digite: **nohup wget -O** código_da_run.nome_espécie.sra link_do_ENA **>/dev/null 2>nome_espécie.err &**

Pressione “Enter”

O computador irá baixar seu dataset. Assim que o processo tiver acabado, podemos abrir as últimas linhas do arquivo “nome_espécie.err” para ver se os dados foram baixados até o fim sem problemas.

Digite: tail nome_espécie.err

As últimas linhas do arquivo serão mostradas na tela, e poderemos ver se o download dos dados foi finalizado sem problemas.

Se a mensagem “Código_da_run.nome_da_espécie.sra saved” aparecer, significa que 100% dos dados foram baixados.

```
gabriel@Leveza: ~/SRA
gabriel@Leveza:~/SRA$ tail Ccostatus.err
10194350K .......... .......... .......... .......... .......... 99%  864K 0s
10194400K .......... .......... .......... .......... .......... 99% 3,36M 0s
10194450K .......... .......... .......... .......... .......... 99% 9,39M 0s
10194500K .......... .......... .......... .......... .......... 99% 1,11M 0s
10194550K .......... .......... .......... .......... .......... 99% 11,0M 0s
10194600K .......... .......... .......... .......... .......... 99% 2,08M 0s
10194650K .......... ..                                          100% 9,87M=99m0s

2018-04-29 21:53:34 (1,68 MB/s) - ‘SRR4342885.Tlaticeps.sra’ saved [10439334134]
```

Agora, precisamos converter os arquivos “.sra” para “.fastq”, que é o formato lido por vários programas de bioinformática. O programa que faz essa conversão está em **/bioinfo/sratoolkit.2.9.0-ubuntu64/bin/fastq-dump.2.9.0** no servidor do nosso laboratório, mas vc mesmo pode fazer o download do pacote de scripts para manipulação de dados do SRA (SRAtoolkit) em: https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/

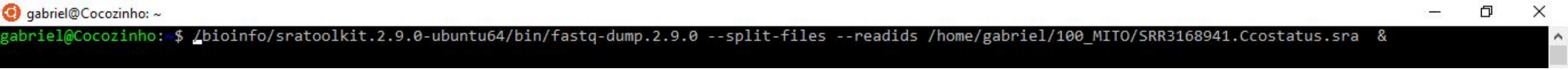
Para dados pareados, devemos usar as opções:

--readids: adiciona código dizendo se a read é forward ou reverse

--split-files: separa os dados em 2 arquivos, um para o forward e outro para o reverse

Digite: caminho/para/fastq-dump –split-files –readids caminho/para/arquivo.sra &

Ex: /bioinfo/sratoolkit.2.9.0-ubuntu64/bin/fastq-dump.2.9.0 --split-files --readids /home/seu_nome_de_usuário/100_MITO/SRR3168941.Ccostatus.sra &



Assim, serão gerados dois arquivos, _1.fastq e _2.fastq. Vamos conferir se eles estão de fato lá usando o comando para listar arquivos.

Digite: ls -ltrh

ls – lista todos os arquivos na pasta

l (long listing)– lista com mais informções

h (human-readable) – tamanho dos arquivos

t (time) – lista arquivos pelo tempo de criação. Dos mais novos para os mais velhos.

r (reverse) – lista pelo reverso do tempo. Ou seja, dos mais velhos para os mais novos.

Com isso, podemos ver que os dois arquivos fastq foram gerados com sucesso.

```
-rw-rw-r-- 1 gabriel gabriel 267M mai 15 14:1  SRR3168941.Ccostatus_2.fastq
-rw-rw-r-- 1 gabriel gabriel 267M mai 15 14:1  SRR3168941.Ccostatus_1.fastq
-rw------- 1 gabriel gabriel 1,2K mai 15 14:1  nohup.out
-rw-rw-r-- 1 gabriel gabriel  16M mai 15 15:40 SRR3168941.Ccostatus.sra
-rw-rw-r-- 1 gabriel gabriel  25K mai 15 15:40 Ccostatus.err
gabriel@Leveza:~/100_MITO$
```

Com as sequências em fastq, podemos finalmente tentar montá-las. Vamos usar um programa chamado NOVOPlasty para isso. Crie uma pasta para a montagem.

Digite: mkdir NOME_DA_ESPÉCIE-NOVOPLASTY

Agora, entre na pasta q vc criou.

Digite: cd nome_da_pasta

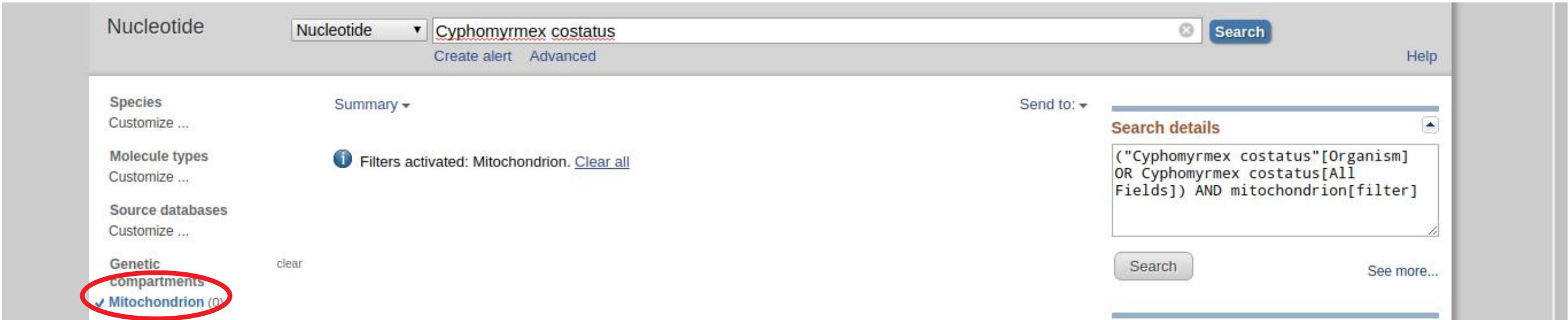
Para iniciar a montagem, o NOVOPlasty precisa de uma “seed”, que nada mais é que uma sequência mitocondrial em formato fasta, que pode ser:

-Uma sequencia mitocondrial da mesma espécie ou de uma espécie bem próxima (mesmo gênero, por exemplo)

-O mitogenoma completo de uma espécie mais distante.

Vamos então procurar por uma seed no genbank. Acesse o site (https://www.ncbi.nlm.nih.gov/genbank/) e procure pela sua espécie, com o filtro para sequências mitocondriais ativado (caso não se lembre como fazer isso, dê uma olhada na parte do tutorial em que conferimos se há mitogenoma completo para a espécie que estamos trabalhando).

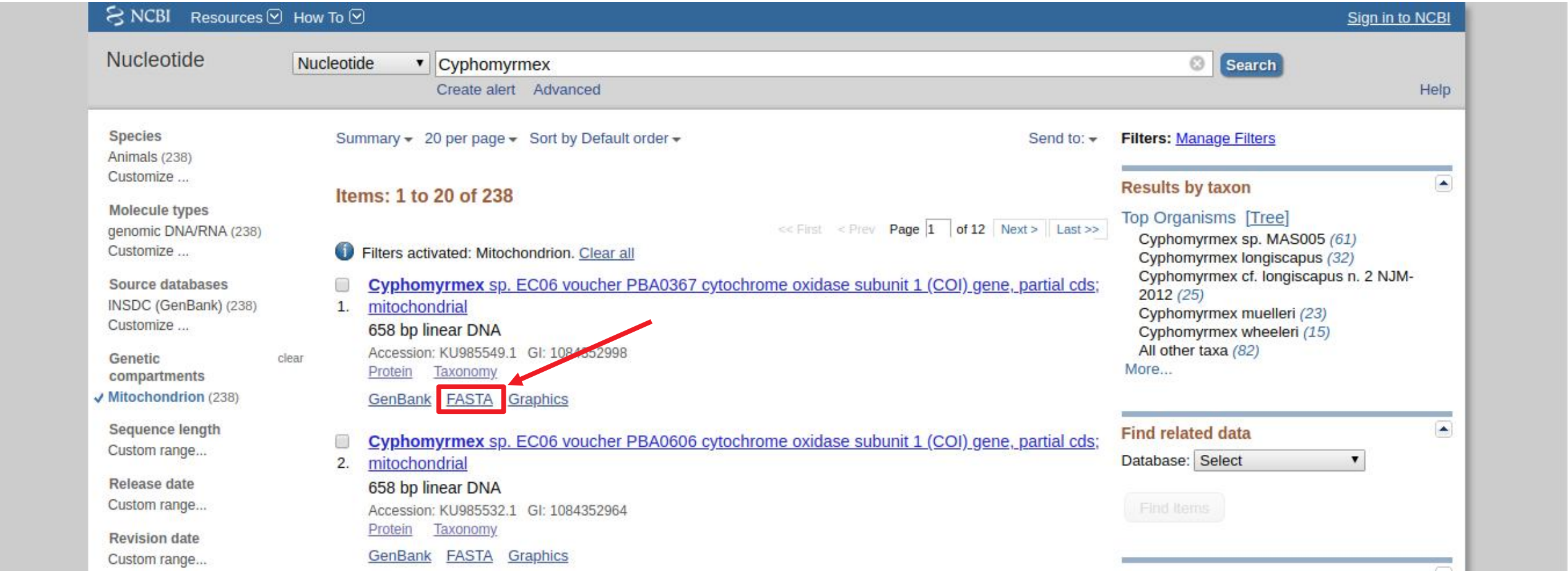
Nesse caso, não há nenhuma sequência mitocondrial para a espécie escolhida. Caso houvesse, poderíamos usá-la como seed.



Em vez de procurar pela mesma espécie, vamos ampliar um pouco e procurar por sequências do mesmo gênero.

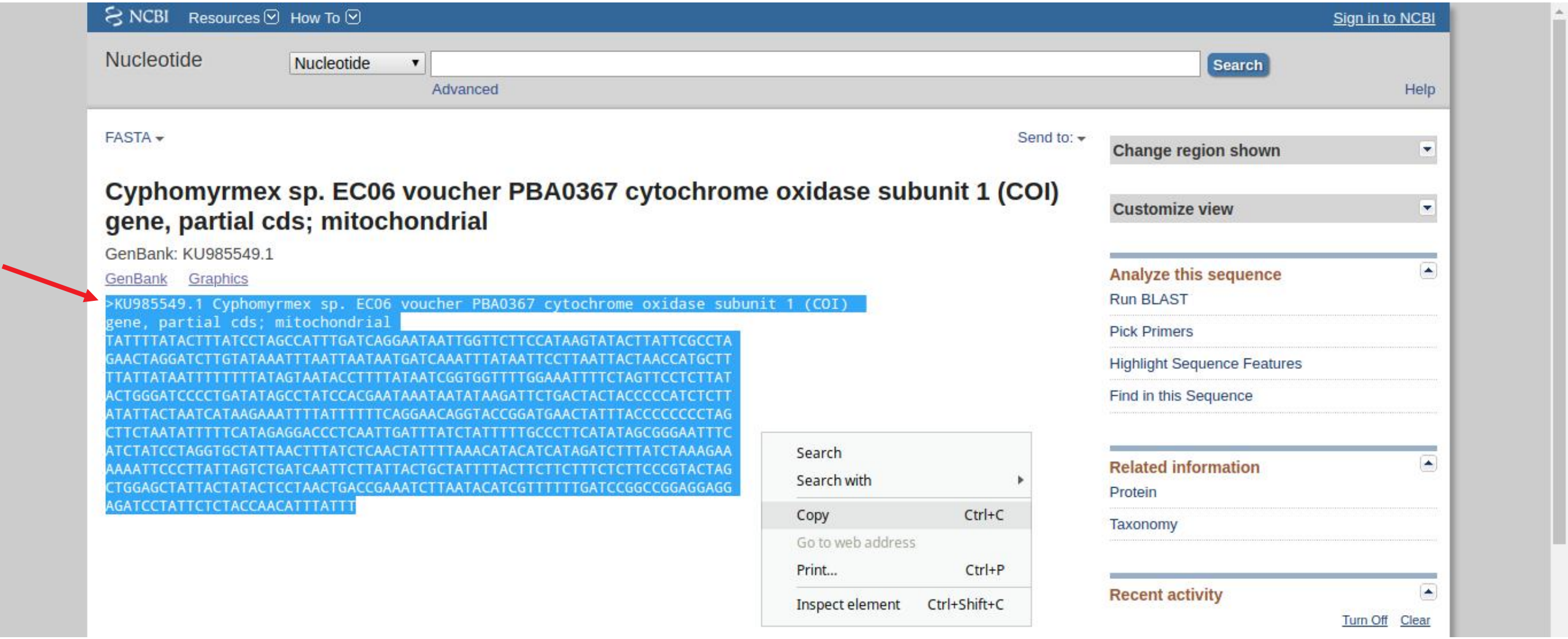
Ao fazer isso, encontramos várias sequências mitocondriais. Vamos usar a primeira que apareceu, que é um fragmento de um gene mitocondrial (COI).

Clique no link “FASTA” abaixo da sequência selecionada como seed.



Você será redirecionado para uma página com a sequência em formato fasta. Selecione e copie ela.

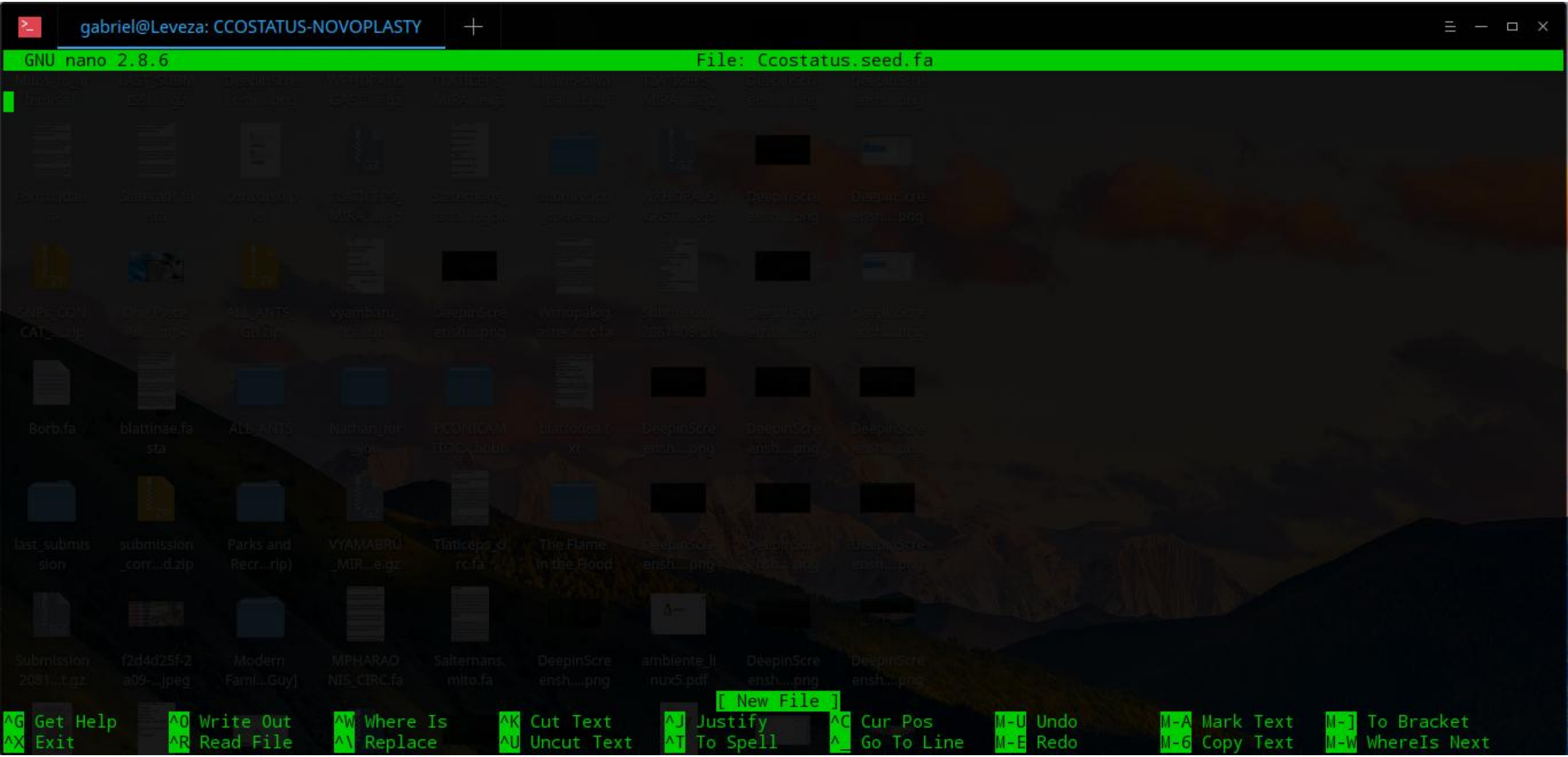
Obs: O arquivo fasta tem duas linhas. A primeira identifica a sequência e começa com um sinal de “>”. A segunda contém a sequẽncia propriamente dita.



Agora, vamos criar um arquivo fasta no servidor. Para isso, usaremos o editor de texto “nano”. Volte ao terminal e, na pasta da montagem, digite:

nano nome_da_espécie.seed.fa

Fazendo isso, você acaba de criar um arquivo fasta (.fa) vazio. Agora, você deve colar a sequência do genbank aqui. Para fazer isso, use Ctrl+Shift+V ou o mouse.



E pronto! Agora temos uma seed que pode ser usada pelo NOVOPlasty para montar sua mitocôndria.

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
GNU nano 2.8.6                    File: Ccostatus.seed.fa                    Modified

>KU985549.1 Cyphomyrmex sp. EC06 voucher PBA0367 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitochondrial
TATTTTATACTTTATCCTAGCCATTTGATCAGGAATAATTGGTTCTTCCATAAGTATACTTATTCGCCTA
GAACTAGGATCTTGTATAAATTTAATTAATAATGATCAAATTTATAATTCCTTAATTACTAACCATGCTT
TTATTATAATTTTTTTTATAGTAATACCTTTTATAATCGGTGGTTTTGGAAATTTTCTAGTTCCTCTTAT
ACTGGGATCCCCTGATATAGCCTATCCACGAATAAATAATATAAGATTCTGACTACTACCCCCATCTCTT
ATATTACTAATCATAAGAAATTTTATTTTTTCAGGAACAGGTACCGGATGAACTATTTACCCCCCCCTAG
CTTCTAATATTTTTCATAGAGGACCCTCAATTGATTTATCTATTTTTGCCCTTCATATAGCGGGAATTTC
ATCTATCCTAGGTGCTATTAACTTTATCTCAACTATTTTAAACATACATCATAGATCTTTATCTAAAGAA
AAAATTCCCTTATTAGTCTGATCAATTCTTATTACTGCTATTTTACTTCTTCTTTCTCTTCCCGTACTAG
CTGGAGCTATTACTATACTCCTAACTGACCGAAATCTTAATACATCGTTTTTTGATCCGGCCGGAGGAGG
AGATCCTATTCTCTACCAACATTTATTT

^G Get Help    ^O Write Out   ^W Where Is    ^K Cut Text    ^J Justify     ^C Cur Pos     M-U Undo       M-A Mark Text  M-] To Bracket
^X Exit        ^R Read File   ^\ Replace     ^U Uncut Text  ^T To Spell    ^_ Go To Line  M-E Redo       M-6 Copy Text  M-W WhereIs Next
```

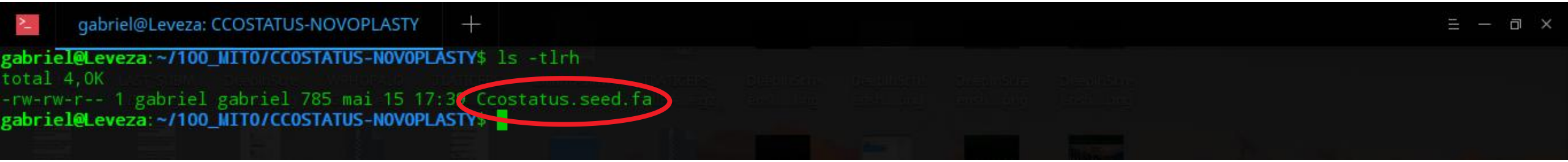
Agora só falta salvar o seu fasta. Para isso, aperte “Ctrl+X”, daí “Y” (para salvar) e “Enter” (para manter o nome do arquivo).

```
^G Get Help    ^O Write Out   ^W Where Is    ^K Cut Text    ^J Justify     ^C Cur Pos     M-U Undo       M-A Mark Text   M-] To Bracket
^X Exit        ^R Read File   ^\ Replace     ^U Uncut Text  ^T To Spell    ^_ Go To Line  M-E Redo       M-6 Copy Text   M-W WhereIs Next
```

```
Save modified buffer?  (Answering "No" will DISCARD changes.)
 Y Yes
 N No           ^C Cancel
```

```
File Name to Write: Ccostatus.seed.fa
^G Get Help    M-D DOS Format    M-A Append    M-B Backup File
^C Cancel      M-M Mac Format    M-P Prepend   ^T To Files
```

Usando “ls -ltrh”, podemos ver que o arquivo foi salvo na pasta da montagem. Vamos para o próximo passo.

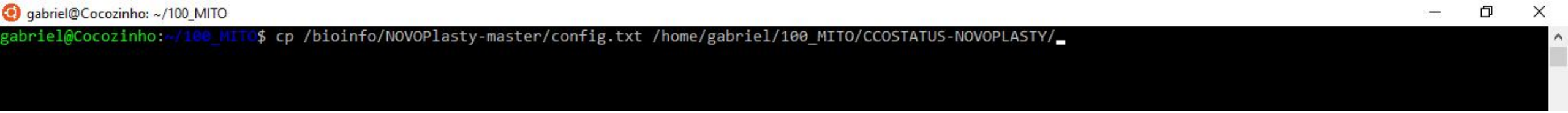


O NOVOPlasty lê um arquivo de configuração (config.txt) que dá para ele todas as informações que ele precisa para rodar. Um molde desse arquivo sempre vem com o programa e iremos copiá-lo para nossa pasta. O caminho completo para esse molde é **/bioinfo/NOVOPlasty-master/config.txt** em nosso servidor.

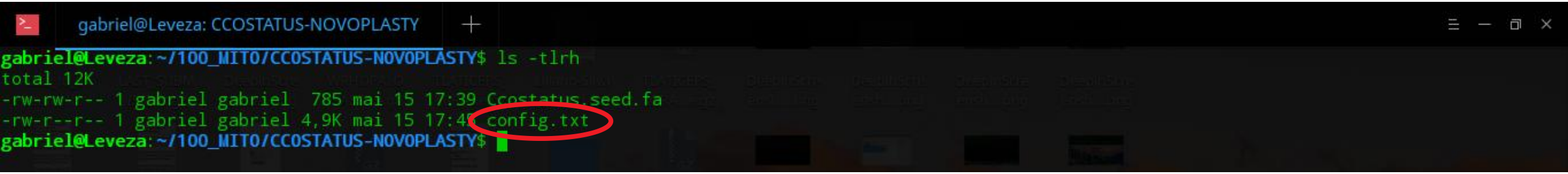
Evidentemente, você também pode baixar o NOVOPlasty em https://github.com/ndierckx/NOVOPlasty e usá-lo localmente.

Para copiá-lo, usaremos o comando “cp” (copy).

Digite: cp /caminho/para/NOVOPlasty-master/config.txt caminho/para/pasta/de/montagem



Ao usar “ls -ltrh”, você verá que o arquivo foi copiado sem problemas.



Agora vamos abrir e editar esse arquivo. Para isso, usaremos o editor de texto “nano”.

Digite: nano caminho/para/config.txt

Isso vai abrir o arquivo de texto, como mostra a imagem:

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
GNU nano 2.8.6                    File: /home/gabriel/100_MITO/CCOSTATUS-NOVOPLASTY/config.txt

Project:
-----------------------
Project name          = Test
Type                  = mito
Genome Range          = 12000-22000
K-mer                 = 39
Max memory            =
Extended log          = 0
Save assembled reads  = no
Seed Input            = Seed.fasta
Reference sequence    = /path/to/reference_file/reference.fasta (optional)
Variance detection    = no
Heteroplasmy          =
Chloroplast sequence  = /path/to/chloroplast_file/chloroplast.fasta (only for "mito_plant" option)

Dataset 1:
-----------------------
Read Length           = 151
Insert size           = 300
Platform              = illumina
Single/Paired         = PE
Combined reads        =
Forward reads         = /path/to/reads/reads_1.fastq
Reverse reads         = /path/to/reads/reads_2.fastq

Optional:
-----------------------
Insert size auto      = yes
Insert Range          = 1.8
Insert Range strict   = 1.3
```

Algumas opções desse arquivo devem ser alteradas:

Project name = Nome_da_espécie

Max memory = 8

Seed Input = caminho/para/seed

Reference sequence = Deve ser apagado e deixado em branco

Chloroplast sequence = Deve ser apagado e deixado em branco

Read length = Tamanho das reads. Deve ser conferido nos arquivos fastq. - **Explicação nos próximos slides**

Insert size = Também dever ser apagado e deixado em branco.

Forward reads = caminho/para/_1.fastq

Reverse reads = caminho/para/_2.fastq

O resto não deve ser alterado.

Para checar o tamanho das reads, basta abrir algum dos fastqs gerados. Vamos checar as 100 primeiras linhas do _1.fastq usando o comando “head”

Digite: head -n 100 caminho/para/arquivo.fastq

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ head -n100 /home/gabriel/100_MITO/SRR3168941.Ccostatus_1.fastq
```

As 100 primeiras linhas do arquivo serão impressas na tela. O tamanho das reads pode ser visto onde estiver escrito “length”.

Nesse caso, o tamanho de cada read é de 100 nucleotídeos. Colocaremos esse valor no config.txt

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ head -n100 /home/gabriel/100_MITO/SRR3168941.Ccostatus_1.fastq
@SRR3168941.1.1 1 length=100
TTNGATACCCAATTTCCACTGTATTTAAAATATGCATACACAAACTATNNTTTTATCATNNNNNNNNNNNNNNNNTATTNGTATCATAAAGTATACGAAT
+SRR3168941.1.1 1 length=100
@C#4ADDFHHHHHJJJJJJJJHIIIJJJJJJIJIIJIJIIIJJJJJJJI##08BGGJIII#################,,=A#,5?BBDEDDDDDDEEFBDDD
@SRR3168941.2.1 2 length=100
TTNATAGAAATTAAGACATACCATAATTGCCGGATGTTACATAGCATANNATTGTATGTGTGTGTGTGTGTGTATGTATATATATATATCTACACACACA
+SRR3168941.2.1 2 length=100
@C#4ADDDHHHHHJJJJJJJJIIIJJJFIHJJJGGIDDFGIJJIJJJJ##07BHFHJJHGCGIJEEEEEEFF@DDDEBDDEEEEEEEEDDDCCCCBBBD
@SRR3168941.3.1 3 length=100
ATNAATTATAAATAATTGGTTAAAAAAATTTGACTAAAAGCAAACAATNNCNATTTAAANNNNNNNNNNNNNNNNTGNCNTATATAGATTTCTTATTTTT
+SRR3168941.3.1 3 length=100
CC#4ADFFHHHHHJIJJJJJJJJJJJJJJJJJJJJJJJJJIIIJJJIIJ##-#-5@E(BE#################,,#,#,5?BDEDDEEDDEDDDEEED
@SRR3168941.4.1 4 length=100
ATTGCATAATTAAAATAATCTCTAGAATATTATATGATATATTTATAGNNTTAATTTCAATTTTAGCAGCATATTTATGTAACAAAAATTTTGCATAGAG
+SRR3168941.4.1 4 length=100
?;@B?DEFHHDFHIIIIIAHHGIIIGHIIIIIEC>>HIIB@FGGFBGC##00?D=BDBBEHGHIIHFGGECHCHG=CDGIIC@@C=ACC=B;@DDBDECC
@SRR3168941.5.1 5 length=100
TTNCCTATACTTGTACCTATACCTAGACGTATTTAGTTTTACGTATTTNNTTTTACGTANNNNNNNNNNNNNNNNTTTTNTATTGCGTTGTACTTAGACG
+SRR3168941.5.1 5 length=100
CC#4ADDFHHHHHJJJJJJJJJJIJJJJJJJEHIJJJJIIJJIJJIJJJJ##08BGGIJHC#################,,=?#,;?BBDDDBDDDEEEDDDDB
@SRR3168941.6.1 6 length=100
TANTTATAATGTATCTAAGATATATTATATTAGCATATAAGAATCGTANNATGAAACGTATCTTACGCATTACAAGGGTAATTCATAATGTATCCAAGTA
+SRR3168941.6.1 6 length=100
CC#4BDFFHHHHHJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJHJ##08DHIJJJIJJJJJJJJGIIJIJHHHHHEDFFFFEEEEDDEEFEDDDDCC
@SRR3168941.7.1 7 length=100
CCNAAAGAACGAAAATCTGCGCTGCAATGTAAAAGAACGAATTATTTCNNGAACTGCTGNNNNNNNNNNNNNNNNGAATNCACCCTGTAGTAAATATAAG
+SRR3168941.7.1 7 length=100
@@#4ADDDHHGGHJJGEHHIJGIIIJJIIEIIGIIIGAHIJGIHGIJA##--@E?EEHF#################,,5<#+8?BDDDDDDDBACEDDCC
@SRR3168941.8.1 8 length=100
ATNCTCATCATGGGAAGAAGACGAGATAACGCCGTGTAATGCTATTCANNAGAATCAGCTAGCCAATCAAATTTTGGTCNTAAATTTTAGTCTTACTGCA
+SRR3168941.8.1 8 length=100
@@#4ABBDFHDFDEHFGGGCCC@:FHEEGHG@FD(??@C@CHGGGGII##-55?EEEHFFC>??@3>@C>@@>ACC;B@#,5<?@A3:B9:>>@>>CCCC
```

Após abrir, editar e salvar o seu config.txt, ele deve ficar parecido com esse:

Select gabriel@Leveza: ~/100_MITO/CCOSTATUS-NOVOPLASTY

```
Project:
-----------------------
Project name          = CCOSTATUS
Type                  = mito
Genome Range          = 12000-22000
K-mer                 = 39
Max memory            = 8
Extended log          = 0
Save assembled reads  = no
Seed Input            = /home/gabriel/100_MITO/CCOSTATUS-NOVOPLASTY/Ccostatus.seed.fa
Reference sequence    =
Variance detection    = no
Heteroplasmy          =
Chloroplast sequence  =

Dataset 1:
-----------------------
Read Length           = 100
Insert size           =
Platform              = illumina
Single/Paired         = PE
Combined reads        =
Forward reads         = /home/gabriel/100_MITO/SRR3168941.Ccostatus_1.fastq
Reverse reads         = /home/gabriel/100_MITO/SRR3168941.Ccostatus_2.fastq

Optional:
-----------------------
Insert size auto      = yes
Insert Range          = 1.8
Insert Range strict   = 1.3
```

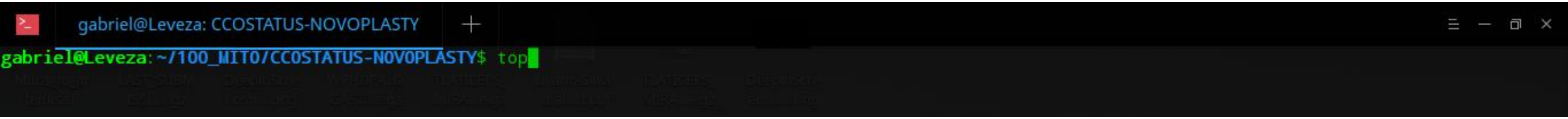
Agora, para finalmente rodar o programa, basta usarmos a seguinte linha de comando:

nohup perl /caminho/para/script/NOVOPlasty -c caminho/para/config.txt >NOVOP.out 2>NOVOP.err &

```
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ nohup perl /bioinfo/NOVOPlasty-master/NOVOPlasty2.7.2.pl  -c /home/gabriel/100_MITO/CCOSTATUS-NOVOPLASTY/config.txt >NOV
OP.out 2>NOVOP.err &
```

Para ver se o programa está rodando normalmente, precisamos usar o comando “top” (não, você não leu errado).

Digite “top” e pressione “Enter”



Tudo o que estiver rodando na máquina irá aparecer, inclusive o NOVOPlasty. Ele aparecerá como “perl”, já que essa é a lingugem de programação na qual ele foi escrito.

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
Tasks: 304 total,   2 running, 302 sleeping,   0 stopped,   0 zombie
%Cpu(s): 11,6 us,  0,5 sy,  0,0 ni, 87,7 id,  0,1 wa,  0,0 hi,  0,1 si,  0,0 st
KiB Mem : 32917216 total,  3306468 free,  2683816 used, 26926932 buff/cache
KiB Swap:  2097148 total,  2097148 free,        0 used. 30028996 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
12271 gabriel   20   0  610208 595348   4312 R  99,7  1,8   0:11.38 perl
 2212 gabi      20   0 2215668 358692 160372 S  37,5  1,1  57:50.62 Web Content
 2140 gabi      20   0 2925516 400868 155612 S   8,0  1,2  10:21.77 firefox
 1263 gabi      20   0 1330792 103448  74068 S   1,0  0,3  18:49.08 Xwayland
 1293 gabi       9 -11 2480132  11876   8696 S   0,7  0,0  11:13.80 pulseaudio
    8 root      20   0       0      0      0 S   0,3  0,0   0:32.04 rcu_sched
 2723 gabi      20   0 2064200 443972 101244 S   0,3  1,3   8:41.41 Web Content
 9700 root      20   0       0      0      0 S   0,3  0,0   0:42.75 kworker/4:0
12151 gabriel   20   0  107696   4560   3508 S   0,3  0,0   0:00.13 sshd
12272 gabriel   20   0   46472   4004   3196 R   0,3  0,0   0:00.02 top
    1 root      20   0  220168   8544   6420 S   0,0  0,0   0:05.85 systemd
    2 root      20   0       0      0      0 S   0,0  0,0   0:00.00 kthreadd
    4 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 kworker/0:0H
    6 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 mm_percpu_wq
    7 root      20   0       0      0      0 S   0,0  0,0   0:00.09 ksoftirqd/0
    9 root      20   0       0      0      0 S   0,0  0,0   0:00.00 rcu_bh
```

Para garantir, você pode apertar “c” durante o top, e o nome do comando rodando irá aparecer.

Caso ele não esteja rodando, abra os arquivos “NOVOP.out” e “NOVOP.err” que foram gerados na pasta da montagem e veja o que pode ter dado errado. Geralmente, são problemas com o config.txt, relativamente fáceis de resolver.

Se estiver rodando, é só esperar a montagem terminar.

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
Tasks: 304 total,   2 running, 302 sleeping,   0 stopped,   0 zombie
%Cpu(s): 11,9 us,  0,3 sy,  0,0 ni, 87,8 id,  0,0 wa,  0,0 hi,  0,1 si,  0,0 st
KiB Mem : 32917216 total,  2677484 free,  3300104 used, 26939628 buff/cache
KiB Swap:  2097148 total,  2097148 free,        0 used. 29413724 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
12271 gabriel   20   0 1199528 1,130g   4340 R 100,0  3,6   5:06.61 perl /bioinfo/NOVOPlasty-master/NOVOPlasty2.6.7.pl -c /home/gabriel/100_MIT0/CC0S+
 2212 gabi      20   0 2227956 379544 160372 S  37,8  1,2  59:42.81 /usr/lib/firefox/firefox -contentproc -childID 1 -isForBrowser -boolPrefs 299:0| +
 2140 gabi      20   0 2927564 401008 155612 S   6,8  1,2  10:41.76 /usr/lib/firefox/firefox -new-window
 1263 gabi      20   0 1330792 103448  74068 S   2,0  0,3  18:52.98 /usr/bin/Xwayland :0 -rootless -terminate -core -listen 4 -listen 5 -displayfd 6
 1293 gabi       9 -11 2480132  11876   8696 S   0,7  0,0  11:15.14 /usr/bin/pulseaudio --start --log-target=syslog
 2723 gabi      20   0 2064200 445908 101244 S   0,7  1,4   8:42.97 /usr/lib/firefox/firefox -contentproc -childID 3 -isForBrowser -boolPrefs 299:0| +
12279 gabriel   20   0   46632   4084   3280 R   0,7  0,0   0:00.02 top
    1 root      20   0  220168   8544   6420 S   0,0  0,0   0:05.86 /sbin/init splash
    2 root      20   0       0      0      0 S   0,0  0,0   0:00.00 [kthreadd]
    4 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 [kworker/0:0H]
    6 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 [mm_percpu_wq]
    7 root      20   0       0      0      0 S   0,0  0,0   0:00.09 [ksoftirqd/0]
    8 root      20   0       0      0      0 S   0,0  0,0   0:32.28 [rcu_sched]
    9 root      20   0       0      0      0 S   0,0  0,0   0:00.00 [rcu_bh]
   10 root      rt   0       0      0      0 S   0,0  0,0   0:00.00 [migration/0]
   11 root      rt   0       0      0      0 S   0,0  0,0   0:00.10 [watchdog/0]
```

A partir daqui, vou usar montagens de outros organismos para exemplificar melhor os três casos possíveis que podem acontecer depois de rodar o NOVOPlasty.

Caso 1: Quando o NOVOPlasty monta o mitogenoma completo

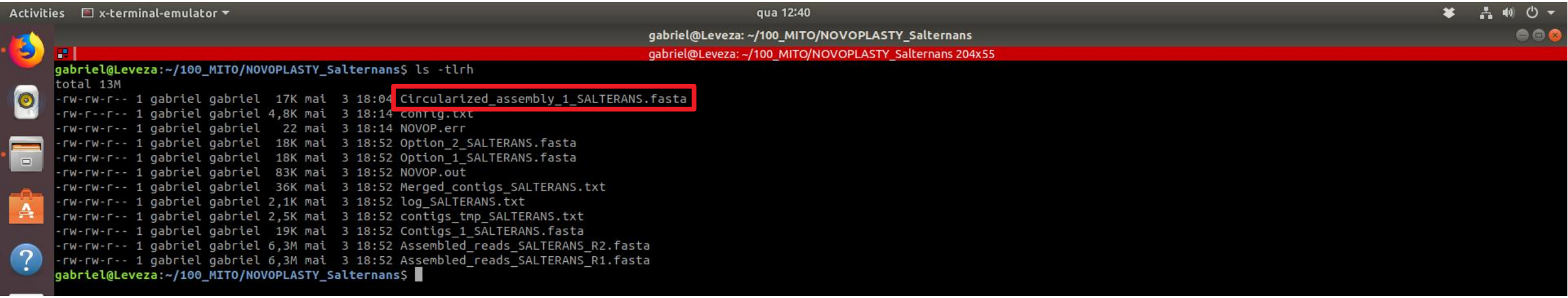
Caso 2: Quando o NOVOPlasty não monta o genoma completo e gera apenas um contig

Caso 3: Quando o NOVOPlasty não monta o genoma completo e gera vários contigs

Caso 1: Quando o NOVOPlasty monta o mitogenoma completo

Nesse caso, ao usar “ls -ltrh”, nós veremos que um arquivo chamado “Circularized_assembly_1_nome_da_espécie.fasta” foi gerado, indicando que a mitocondria inteira foi montada.

Se esse arquivo não for gerado, vc deve ver se sua montagem se encaixa no Caso 2 ou 3.

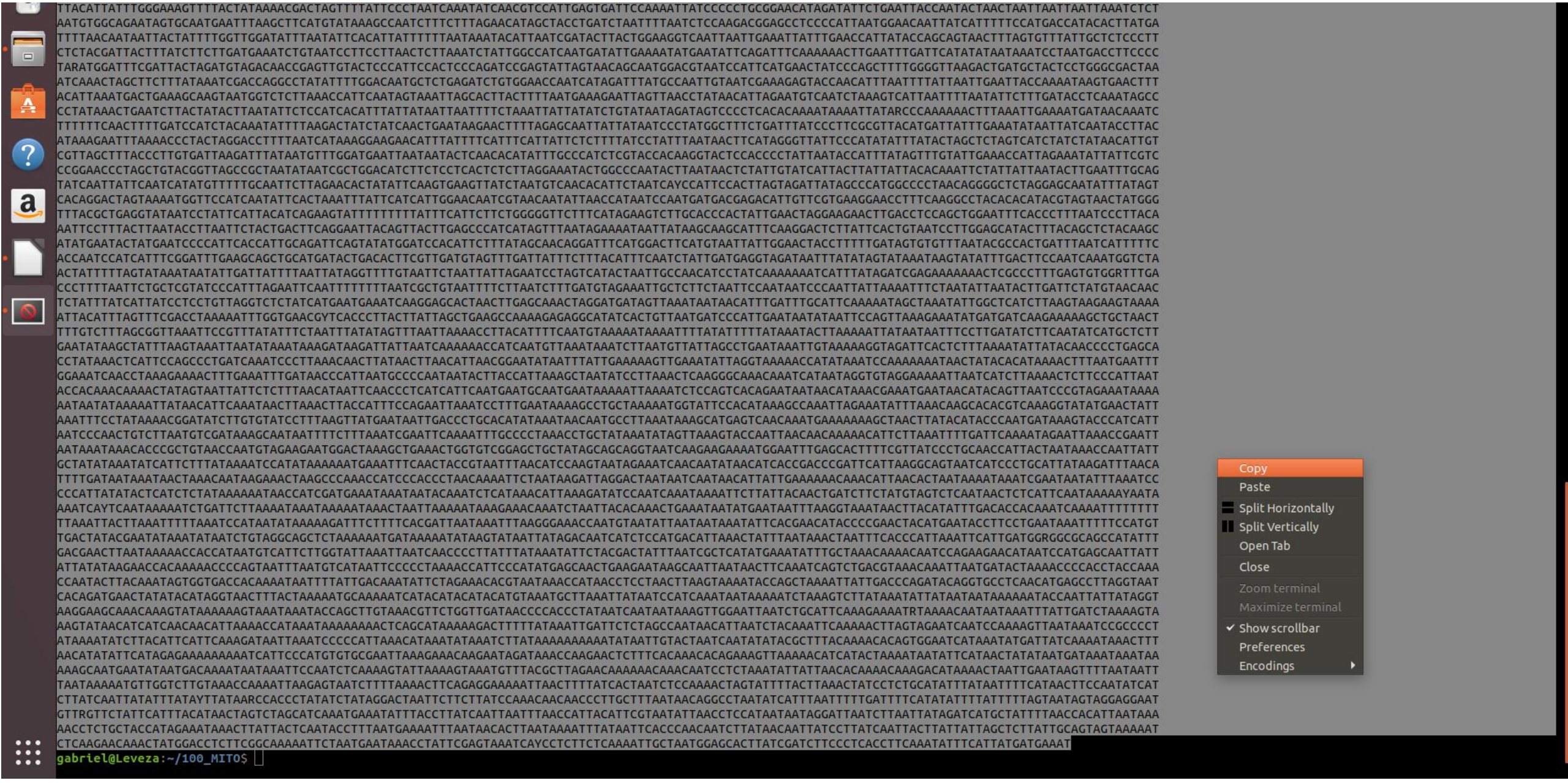


Agora vamos usar o comando “cat”, que irá imprimir o conteúdo do arquivo “Circularized_assembly” na tela:

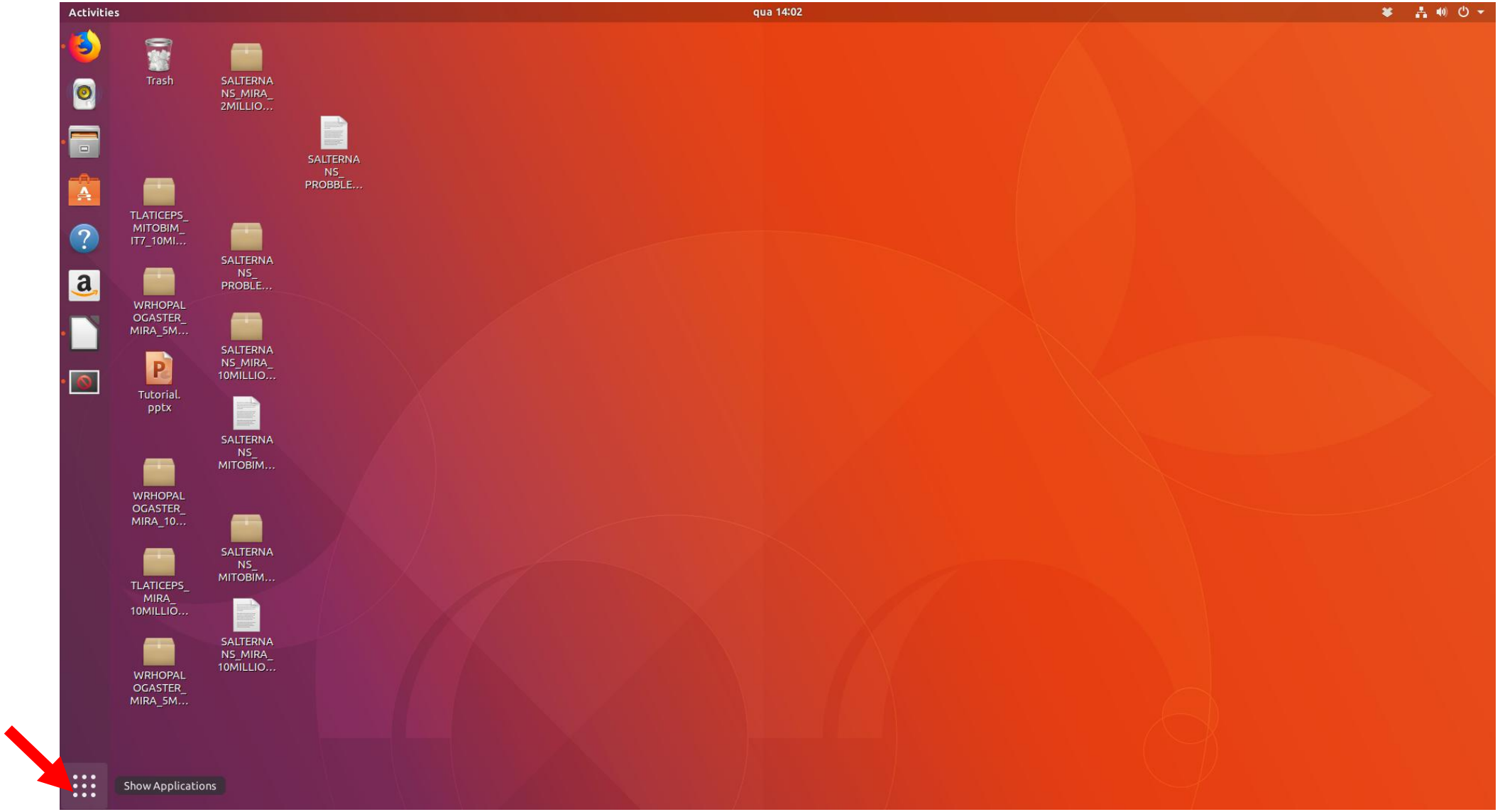
Digite: cat /caminho/para/Circularized_assembly_1_nome_da_espécie.fasta

E a sequencia mitocondrial completa irá aparecer na tela.

Agora, selecione a sequência até o final com o mouse e a copie.

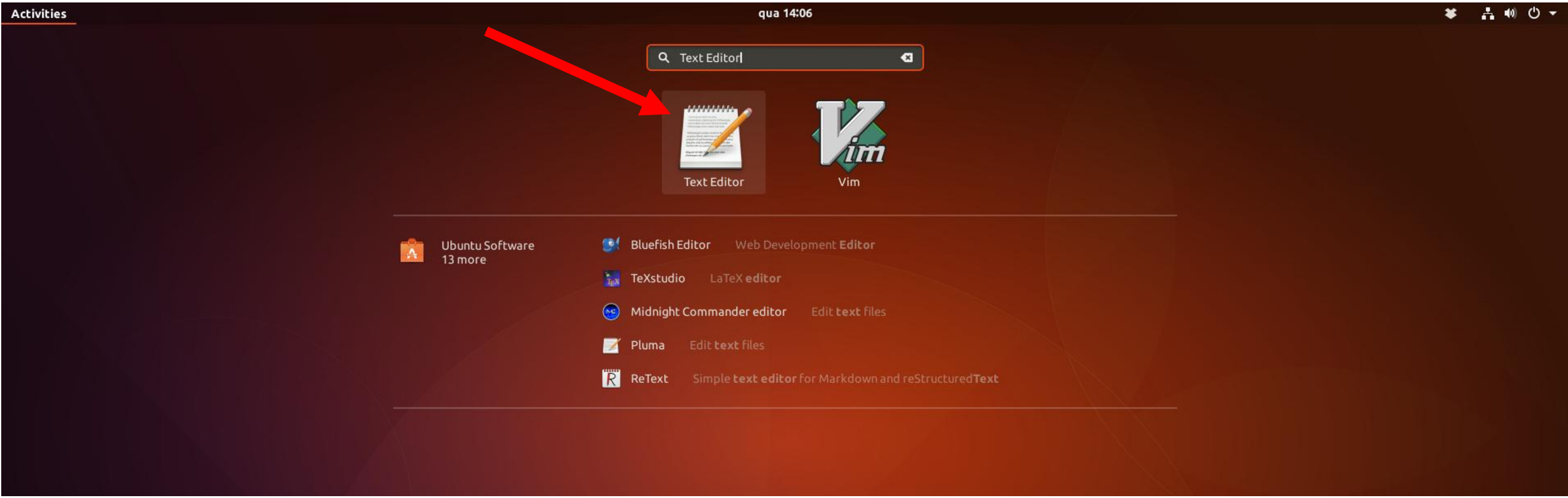


Agora, abra o editor de texto do seu computador. No caso do Ubuntu, basta clicar em Show Applications, na parte inferior esquerda da tela, e depois pesquisar por “Text Editor”.

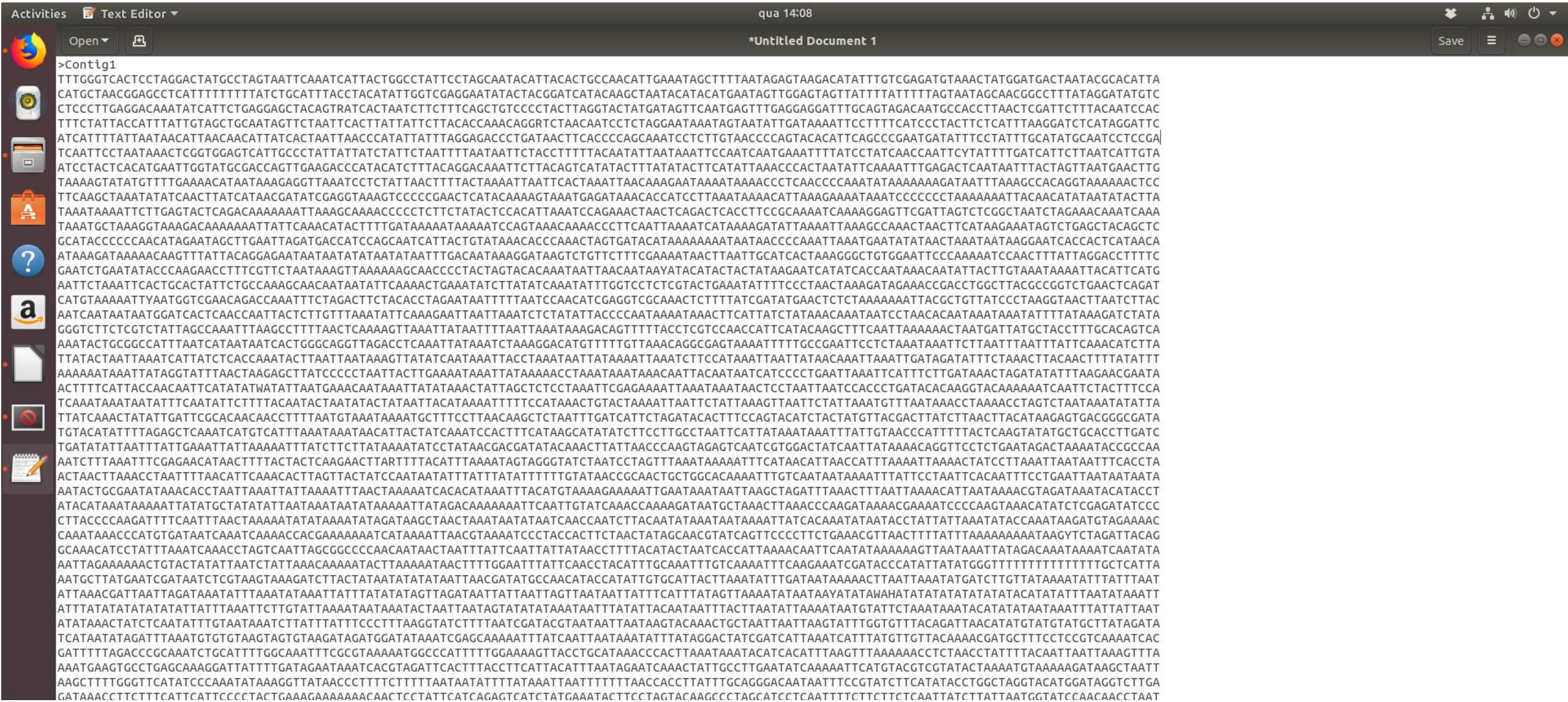


Agora, abra o editor de texto do seu computador. No caso do Ubuntu, basta clicar em Show Applications, na parte inferior esquerda da tela, e depois pesquisar por “Text Editor”.

Obs: Qualquer editor de texto pode ser usado para essa etapa.



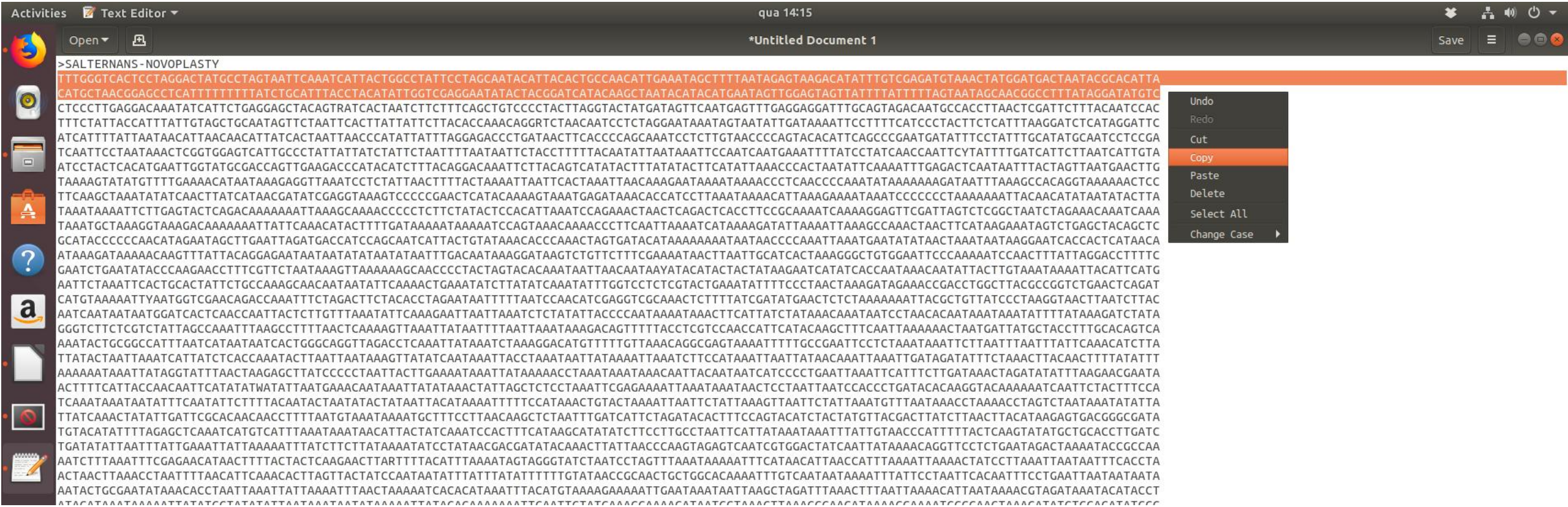
Abra o editor de texto e cole o resultado da montagem.




```
>Contig1
TTTGGGTCACTCCTAGGACTATGCCTAGTAATTCAAATCATTACTGGCCTATTCCTAGCAATACATTACACTGCCAACATTGAAATAGCTTTTAATAGAGTAAGACATATTTGTCGAGATGTAAACTATGGATGACTAATACGCACATTA
CATGCTAACGGAGCCTCATTTTTTTTTATCTGCATTTACCTACATATTGGTCGAGGAATATACTACGGATCATACAAGCTAATACATACATGAATAGTTGGAGTAGTTATTTTATTTTTAGTAATAGCAACGGCCTTTATAGGATATGTC
CTCCCTTGAGGACAAATATCATTCTGAGGAGCTACAGTRATCACTAATCTTCTTTCAGCTGTCCCCTACTTAGGTACTATGATAGTTCAATGAGTTTGAGGAGGATTTGCAGTAGACAATGCCACCTTAACTCGATTCTTTACAATCCAC
TTTCTATTACCATTTATTGTAGCTGCAATAGTTCTAATTCACTTATTATTCTTACACCAAACAGGRTCTAACAATCCTCTAGGAATAAATAGTAATATTGATAAAATTCCTTTTCATCCCTACTTCTCATTTAAGGATCTCATAGGATTC
ATCATTTTATTAATAACATTAACAACATTATCACTAATTAACCCATATTATTTAGGAGACCCTGATAACTTCACCCCAGCAAATCCTCTTGTAACCCCAGTACACATTCAGCCCGAATGATATTTCCTATTTGCATATGCAATCCTCCGA
TCAATTCCTAATAAACTCGGTGGAGTCATTGCCCTATTATTATCTATTCTAATTTTAATAATTCTACCTTTTTACAATATTAATAAATTCCAATCAATGAAATTTTATCCTATCAACCAATTCYTATTTTGATCATTCTTAATCATTGTA
ATCCTACTCACATGAATTGGTATGCGACCAGTTGAAGACCCATACATCTTTACAGGACAAATTCTTACAGTCATATACTTTATATACTTCATATTAAACCCACTAATATTCAAAATTTGAGACTCAATAATTTACTAGTTAATGAACTTG
TAAAAGTATATGTTTTGAAAACATAATAAAGAGGTTAAATCCTCTATTAACTTTTACTAAAATTAATTCACTAAATTAACAAAGAATAAAATAAAACCCTCAACCCCAAATATAAAAAAAGATAATTTAAAGCCACAGGTAAAAAACTCC
TTCAAGCTAAATATATCAACTTATCATAACGATATCGAGGTAAAGTCCCCCGAACTCATACAAAAGTAAATGAGATAAACACCATCCTTAAATAAAACATTAAAGAAAATAAATCCCCCCCTAAAAAAATTACAACATATAATATACTTA
TAAATAAAATTCTTGAGTACTCAGACAAAAAAATTAAAGCAAAACCCCCTCTTCTATACTCCACATTAAATCCAGAAACTAACTCAGACTCACCTTCCGCAAAATCAAAAGGAGTTCGATTAGTCTCGGCTAATCTAGAAACAAATCAAA
TAAATGCTAAAGGTAAAGACAAAAAAATTATTCAAACATACTTTTGATAAAAATAAAAATCCAGTAAACAAAACCCTTCAATTAAAATCATAAAAGATATTAAAATTAAAGCCAAACTAACTTCATAAGAAATAGTCTGAGCTACAGCTC
GCATACCCCCCAACATAGAATAGCTTGAATTAGATGACCATCCAGCAATCATTACTGTATAAACACCCAAACTAGTGATACATAAAAAAAATAATAACCCCAAATTAAATGAATATATAACTAAATAATAAGGAATCACCACTCATAACA
ATAAAGATAAAAACAAGTTTATTACAGGAGAATAATAATATATAATATAATTTGACAATAAAGGATAAGTCTGTTCTTTCGAAAATAACTTAATTGCATCACTAAAGGGCTGTGGAATTCCCAAAAATCCAACTTTATTAGGACCTTTTC
GAATCTGAATATACCCAAGAACCTTTCGTTCTAATAAAGTTAAAAAAGCAACCCCTACTAGTACACAAATAATTAACAATAAYATACATACTACTATAAGAATCATATCACCAATAAACAATATTACTTGTAAATAAAATTACATTCATG
AATTCTAAATTCACTGCACTATTCTGCCAAAGCAACAATAATATTCAAAACTGAAATATCTTATATCAAATATTTGGTCCTCTCGTACTGAAATATTTTCCCTAACTAAAGATAGAAACCGACCTGGCTTACGCCGGTCTGAACTCAGAT
CATGTAAAAATTYAATGGTCGAACAGACCAAATTTCTAGACTTCTACACCTAGAATAATTTTTAATCCAACATCGAGGTCGCAAACTCTTTTATCGATATGAACTCTCTAAAAAAATTACGCTGTTATCCCTAAGGTAACTTAATCTTAC
AATCAATAATAATGGATCACTCAACCAATTACTCTTGTTTAAATATTCAAAGAATTAATTAAATCTCTATATTACCCCAATAAAATAAACTTCATTATCTATAAACAAATAATCCTAACACAATAAATAAATATTTTATAAAGATCTATA
GGGTCTTCTCGTCTATTAGCCAAATTTAAGCCTTTTAACTCAAAAGTTAAATTATAATTTTAATTAAATAAAGACAGTTTTTACCTCGTCCAACCATTCATACAAGCTTTCAATTAAAAAACTAATGATTATGCTACCTTTGCACAGTCA
AAATACTGCGGCCATTTAATCATAATAATCACTGGGCAGGTTAGACCTCAAATTATAAATCTAAAGGACATGTTTTTGTTAAACAGGCGAGTAAAATTTTTGCCGAATTCCTCTAAATAAATTCTTAATTTAATTTATTCAAACATCTTA
TTATACTAATTAAATCATTATCTCACCAAATACTTAATTAATAAAGTTATATCAATAAATTACCTAAATAATTATAAAATTAAATCTTCCATAAATTAATTATAACAAATTAAATTGATAGATATTTCTAAACTTACAACTTTTATATTT
AAAAAATAAATTATAGGTATTTAACTAAGAGCTTATCCCCCTAATTACTTGAAAATAAATTATAAAAACCTAAATAAATAAACAATTACAATAATCATCCCCTGAATTAAATTCATTTCTTGATAAACTAGATATATTTAAGAACGAATA
ACTTTTCATTACCAACAATTCATATATWATATTAATGAAACAATAAATTATATAAACTATTAGCTCTCCTAAATTCGAGAAAATTAAATAAATAACTCCTAATTAATCCACCCTGATACACAAGGTACAAAAAATCAATTCTACTTTCCA
TCAAATAAATAATATTTCAATATTCTTTTACAATACTAATATACTATAATTACATAAAATTTTTCCATAAACTGTACTAAAATTAATTCTATTAAAGTTAATTCTATTAAATGTTTAATAAACCTAAAACCTAGTCTAATAAATATATTA
TTATCAAACTATATTGATTCGCACAACAACCTTTTAATGTAAATAAAATGCTTTCCTTAACAAGCTCTAATTTGATCATTCTAGATACACTTTCCAGTACATCTACTATGTTACGACTTATCTTAACTTACATAAGAGTGACGGGCGATA
TGTACATATTTTAGAGCTCAAATCATGTCATTTAAATAAATAACATTACTATCAAATCCACTTTCATAAGCATATATCTTCCTTGCCTAATTCATTATAAATAAATTTATTGTAACCCATTTTTACTCAAGTATATGCTGCACCTTGATC
TGATATATTAATTTATTGAAATTATTAAAAATTTATCTTCTTATAAAATATCCTATAACGACGATATACAAACTTATTAACCCAAGTAGAGTCAATCGTGGACTATCAATTATAAAACAGGTTCCTCTGAATAGACTAAAATACCGCCAA
AATCTTTAAATTTCGAGAACATAACTTTTACTACTCAAGAACTTARTTTTACATTTAAAATAGTAGGGTATCTAATCCTAGTTTAAATAAAAATTTCATAACATTAACCATTTAAAATTAAAACTATCCTTAAATTAATAATTTCACCTA
ACTAACTTAAACCTAATTTTAACATTCAAACACTTAGTTACTATCCAATAATATTTATTTATATTTTTTGTATAACCGCAACTGCTGGCACAAAATTTGTCAATAATAAAATTTATTCCTAATTCACAATTTCCTGAATTAATAATAATA
AATACTGCGAATATAAACACCTAATTAAATTATTAAAATTTAACTAAAAATCACACATAAATTTACATGTAAAAGAAAAATTGAATAAATAATTAAGCTAGATTTAAACTTTAATTAAAACATTAATAAAACGTAGATAAATACATACCT
ATACATAAATAAAAATTATATGCTATATATTAATAAATAATATAAAAATTATAGACAAAAAAATTCAATTGTATCAAACCAAAAGATAATGCTAAACTTAAACCCAAGATAAAACGAAAATCCCCAAGTAAACATATCTCGAGATATCCC
CTTACCCCAAGATTTTCAATTTAACTAAAAATATATAAAATATAGATAAGCTAACTAAATAATATAATCAACCAATCTTACAATATAAATAATAAAATTATCACAAATATAATACCTATTATTAAATATACCAAATAAGATGTAGAAAAC
CAAATAAACCCATGTGATAATCAAATCAAAACCACGAAAAAAATCATAAAATTAACGTAAAATCCCTACCACTTCTAACTATAGCAACGTATCAGTTCCCCTTCTGAAACGTTAACTTTTATTTAAAAAAAAATAAGYTCTAGATTACAG
GCAAACATCCTATTTAAATCAAACCTAGTCAATTAGCGGCCCCAACAATAACTAATTTATTCAATTATTATAACCTTTTACATACTAATCACCATTAAAACAATTCAATATAAAAAAGTTAATAAATTATAGACAAATAAAATCAATATA
AATTAGAAAAAACTGTACTATATTAATCTATTAAACAAAAATACTTAAAAATAACTTTTGGAATTTATTCAACCTACATTTGCAAATTTGTCAAAATTTCAAGAAATCGATACCCATATTATATGGGTTTTTTTTTTTTTTTGCTCATTA
AATGCTTATGAATCGATAATCTCGTAAGTAAAGATCTTACTATAATATATATAATTAACGATATGCCAACATACCATATTGTGCATTACTTAAATATTTGATAATAAAAACTTAATTAAATATGATCTTGTTATAAAATATTTATTTAAT
ATTAAACGATTAATTAGATAAATATTTAAATATAAATTATTTATATATAGTTAGATAATTATTAATTAGTTAATAATTATTTCATTTATAGTTAAAATATAATAAYATATAWAHATATATATATATATATACATATATTTAATATAAATT
ATTTATATATATATATATTATTTAAATTCTTGTATTAAAATAATAAATACTAATTAATAGTATATATAAATAATTTATATTACAATAATTTACTTAATATTAAAATAATGTATTCTAAATAAATACATATATAATAAATTTATTATTAAT
ATATAAACTATCTCAATATTTGTAATAAATCTTATTTATTTCCCTTTAAGGTATCTTTTAATCGATACGTAATAATTAATAAGTACAAACTGCTAATTAATTAAGTATTTGGTGTTTACAGATTAACATATGTATGTATGCTTATAGATA
TCATAATATAGATTTAAATGTGTGTAAGTAGTGTAAGATAGATGGATATAAATCGAGCAAAAATTTATCAATTAATAAATATTTATAGGACTATCGATCATTAAATCATTTATGTTGTTACAAAACGATGCTTTCCTCCGTCAAAATCAC
GATTTTTAGACCCGCAAATCTGCATTTTGGCAAATTTCGCGTAAAAATGGCCCATTTTTGGAAAAGTTACCTGCATAAACCCACTTAAATAAATACATCACATTTAAGTTTAAAAAACCTCTAACCTATTTTACAATTAATTAAAGTTTA
AAATGAAGTGCCTGAGCAAAGGATTATTTTGATAGAATAAATCACGTAGATTCACTTTACCTTCATTACATTTAATAGAATCAAACTATTGCCTTGAATATCAAAAATTCATGTACGTCGTATACTAAAATGTAAAAAGATAAGCTAATT
AAGCTTTTGGGTTCATATCCCAAATATAAAGGTTATAACCCTTTTCTTTTTAATAATATTTTATAAATTAATTTTTTTAACCACCTTATTTGCAGGGACAATAATTTCCGTATCTTCATATACCTGGCTAGGTACATGGATAGGTCTTGA
GATAAACCTTCTTTCATTCATTCCCCTACTGAAAGAAAAAAACAACTCCTATTCATCAGAGTCATCTATGAAATACTTCCTAGTACAAGCCCTAGCATCCTCAATTTTCTTCTTCTCAATTATCTTATTAATGGTATCCAACAACCTAAT
```

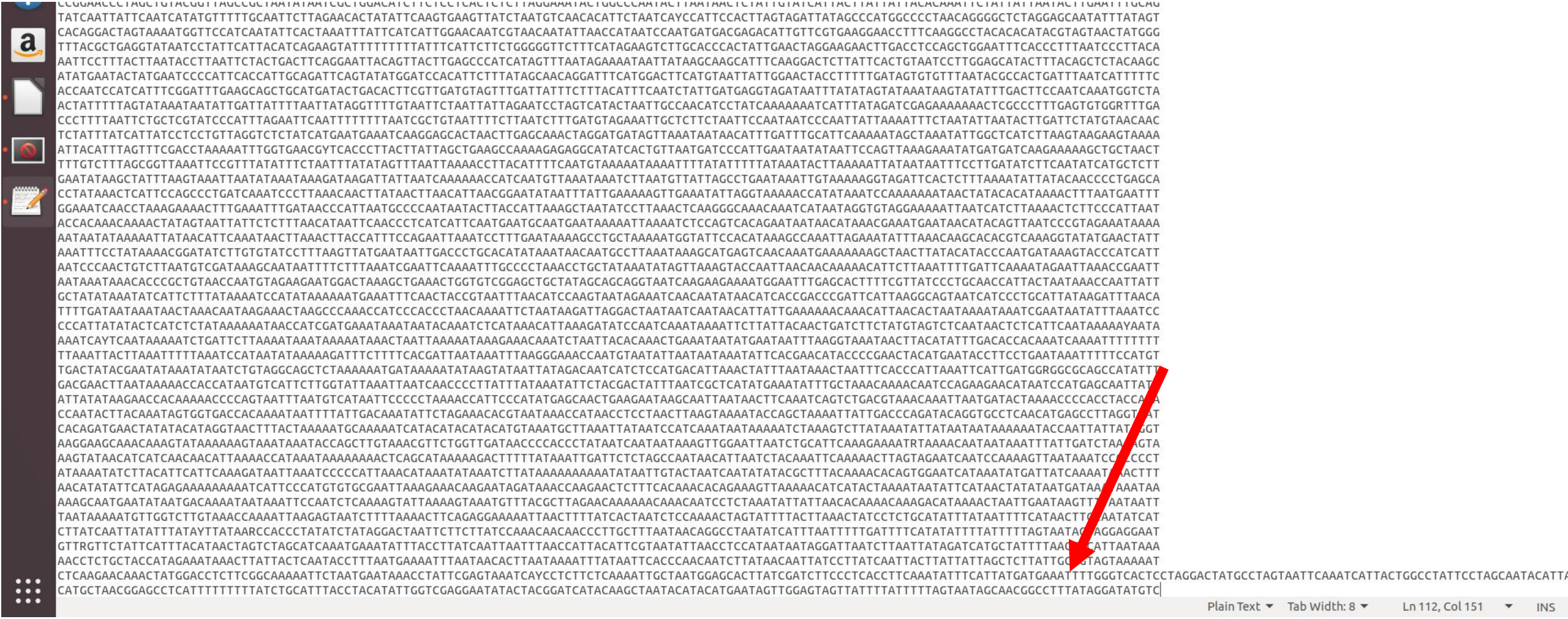
Edite a primeira linha do fasta. Substitua “>Contig1” por “>Nome_da_espécie-NOVOPLASTY”

Selecione as duas primeiras linhas de nucleotídeos do arquivo e as copie.

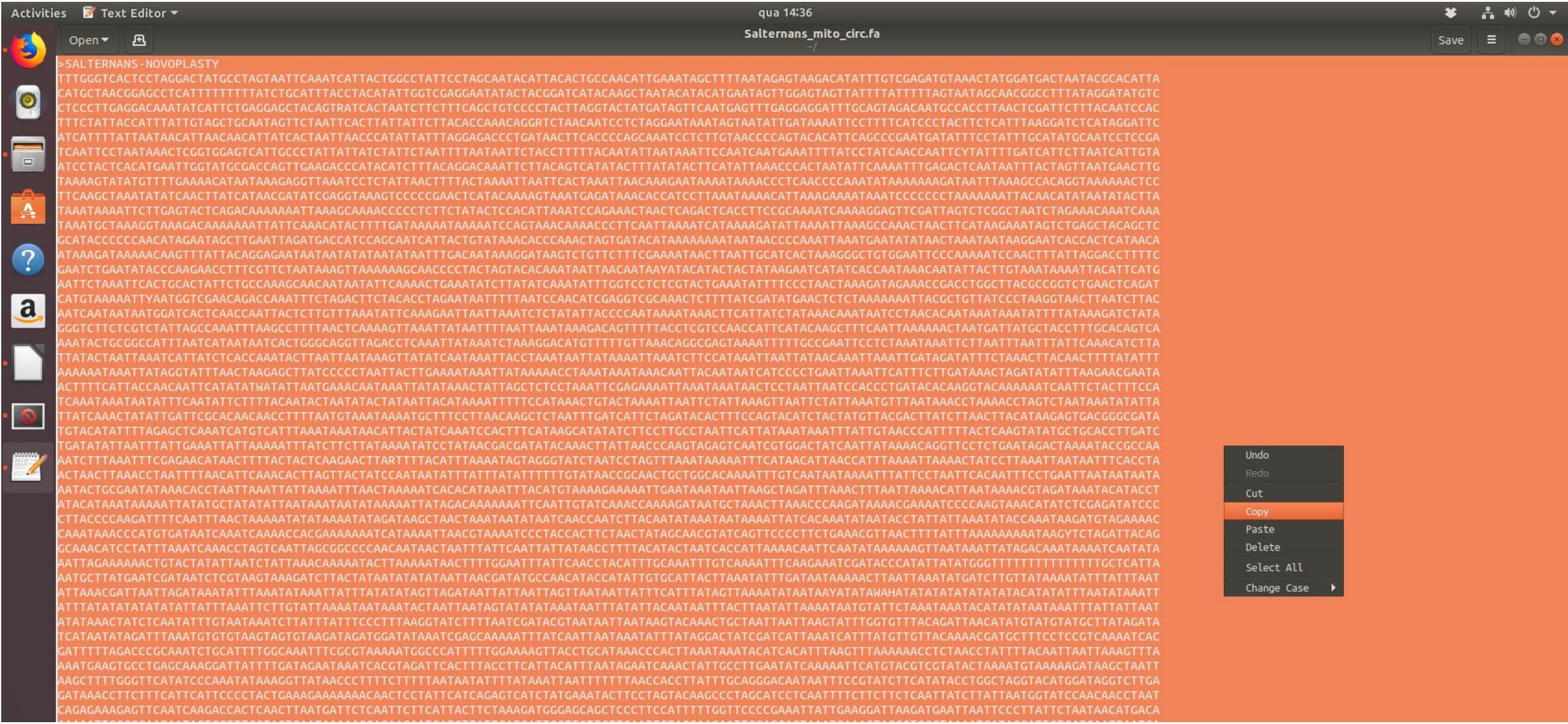


Agora, vá até o fim do arquivo (após o último nucleotídeo) e as cole. Por fim, salve o arquivo com o nome “nome_da_espécie.mito.circ.fa”. Isso é importante para checarmos depois se o mitogenoma está de fato circularizado.

Agora, vá até o fim do arquivo (após o último nucleotídeo) e as cole. Por fim, salve o arquivo com o nome “nome_da_espécie.mito.circ.fa”. Isso é importante para checarmos depois se o mitogenoma está de fato circularizado.



Agora, copie toda a sequencia desse novo arquivo (Ctrl+a para selecionar tudo e Ctrl+c para copiar)



Voltando ao terminal, vamos criar um arquivo fasta na pasta da montagem NOVOPlasty usando o nano. Cole a sequencia (Ctrl+Shift+C) e salve.

nano nome_da_espécie.mito.circ.fa

Caso 2: Quando o NOVOPlasty não monta o genoma completo e gera apenas um contig

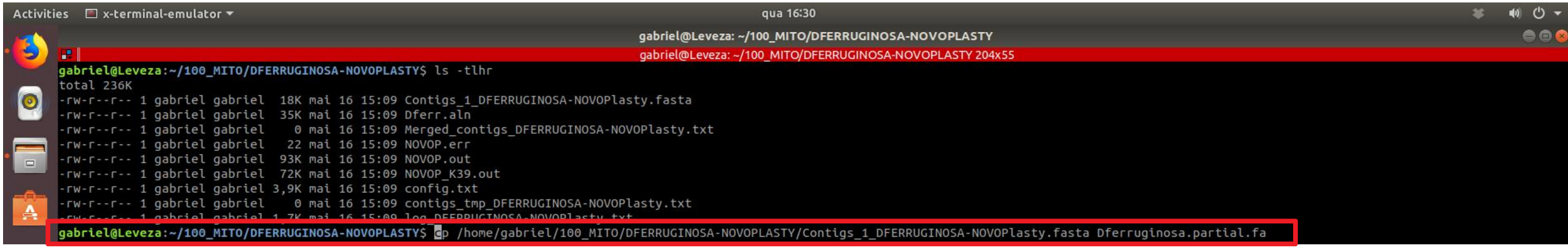
Depois que a montagem terminar, use “ls -ltrh” para ver os arquivos gerados.

Nesse caso, a montagem não obteve o mitogenoma completo, já que não gerou um arquivo chamado “Circulared_assembly”.

Vamos fazer uma cópia do arquivo “Contigs_1_nome_da_espécie.fasta” usando “cp”.

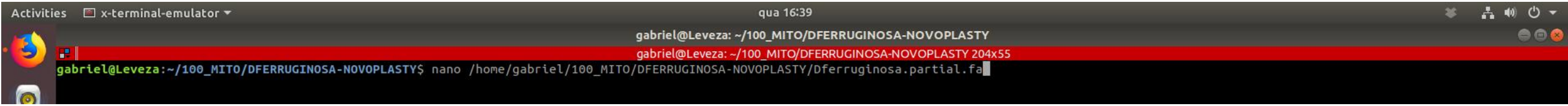
Essa cópia se chamara “Nome_da_espécie.partial.fa”

cp /caminho/para/Contigs_1.fasta Nome_da_espécie.partial.fa

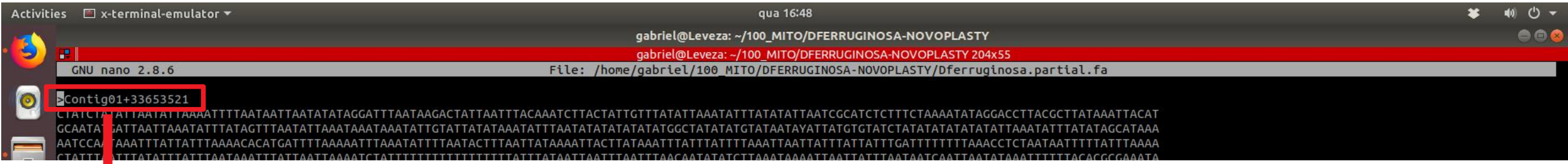


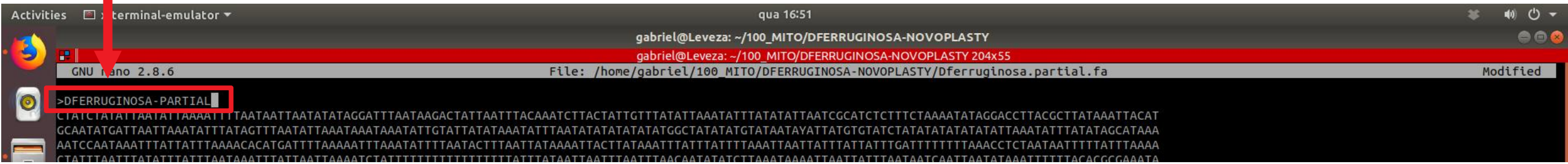
Agora, abra o arquivo que você criou usando o nano.

nano /caminho/para/Nome_da_espécie.partial.fa



E, por último, edite a primeira linha do fasta. Substitua por “>Nome_da_espécie-PARTIAL” e salve o arquivo (“Ctrl+X”, depois “Y”, depois “Enter”)





Caso 3: Quando o NOVOPlasty não monta o genoma completo e gera vários contigs

Depois que a montagem terminar, use “ls -ltrh” para ver os arquivos gerados.

Nesse caso, a montagem não obteve o mitogenoma completo, já que não gerou um arquivo chamado “Circularized_assembly”.

Em vez disso, foi gerado um arquivo “Contigs_1_Nome_da_espécie.fasta”.

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ ls -tlrh
total 456K
-rw-rw-r-- 1 gabriel gabriel  785 mai 15 17:39 Ccostatus.seed.fa
-rw-r--r-- 1 gabriel gabriel 4,9K mai 15 19:10 config.txt
-rw-rw-r-- 1 gabriel gabriel   22 mai 15 19:20 NOVOP.err
-rw-rw-r-- 1 gabriel gabriel  45K mai 15 19:28 NOVOP.out
-rw-rw-r-- 1 gabriel gabriel 2,2K mai 15 19:28 log_CCOSTATUS.txt
-rw-rw-r-- 1 gabriel gabriel 8,0K mai 15 19:28 contigs_tmp_CCOSTATUS.txt
-rw-rw-r-- 1 gabriel gabriel 5,8K mai 15 19:28 Contigs_1_CCOSTATUS.fasta
-rw-rw-r-- 1 gabriel gabriel 177K mai 15 19:28 Assembled_reads_CCOSTATUS_R2.fasta
-rw-rw-r-- 1 gabriel gabriel 177K mai 15 19:28 Assembled_reads_CCOSTATUS_R1.fasta
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$
```

Agora vamos usar o comando “cat”, que irá imprimir o conteúdo do arquivo “Contigs_1” na tela:

Digite: cat /caminho/para/Contigs_1_nome_da_espécie.fasta

E as sequencias mitocondriais irão aparecer na tela. Podemos ver que mais de um contig foi gerado.

```
gabriel@Leveza: CCOSTATUS-NOVOPLASTY
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ cat /home/gabriel/100_MITO/CCOSTATUS-NOVOPLASTY/Contigs_1_CCOSTATUS.fasta
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
```

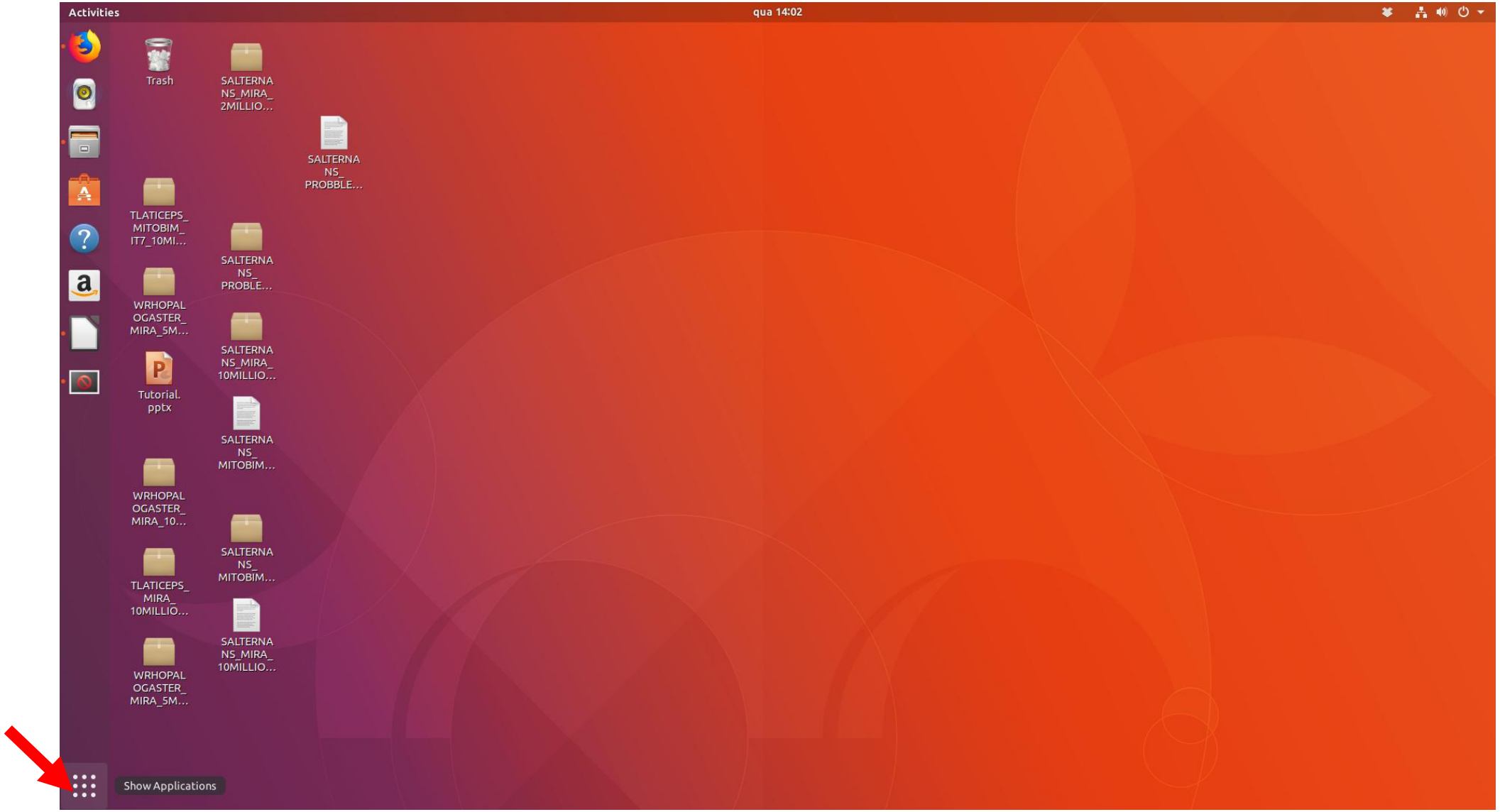
Com o mouse, selecione todos os contigs gerados e os copie.

gabriel@Leveza: CCOSTATUS-NOVOPLASTY

```
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ cat /home/gabriel/100_MITO/CCOSTATUS-NOVOPLASTY/Contigs_1_CCOSTATUS.fasta
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTAC
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTT
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCT
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTT
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTA
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCAC
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTT
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTT
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCC
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTG
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTGTTTCTTTTTTTGAACCTACTTTCACTCTTCCCTAGC
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCA
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATA
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTT
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTAC
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
```

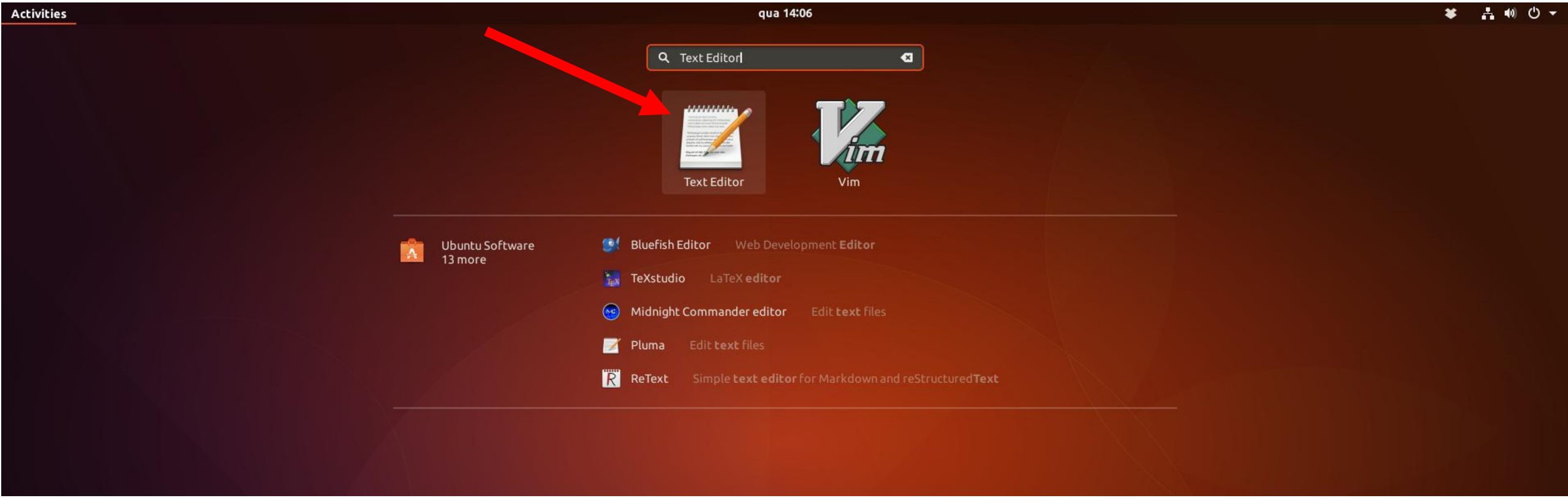
Copy
Paste
Horizontal split
Vertical split
Close window
New workspace
Fullscreen
Search
Google
Bing
Rename title
Encoding
Custom commands
Remote management
Settings

Agora, abra o editor de texto do seu computador. No caso do Ubuntu, basta clicar em Show Applications, na parte inferior esquerda da tela, e depois pesquisar por “Text Editor”.



Agora, abra o editor de texto do seu computador. No caso do Ubuntu, basta clicar em Show Applications, na parte inferior esquerda da tela, e depois pesquisar por “Text Editor”.

Obs: Qualquer editor de texto pode ser usado para essa etapa.



Abra o editor de texto de sua escolha e cole o resultado da montagem.

Open ▾ *Untitled Document 1 Save

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
```

Agora, iremos ver se as extremidades (primeira e última linhas) dos contigs gerados aparecem em outros contigs, para assim fundi-los e obter uma sequência maior. Para isso, vamos usar a ferramenta de busca do editor de texto.

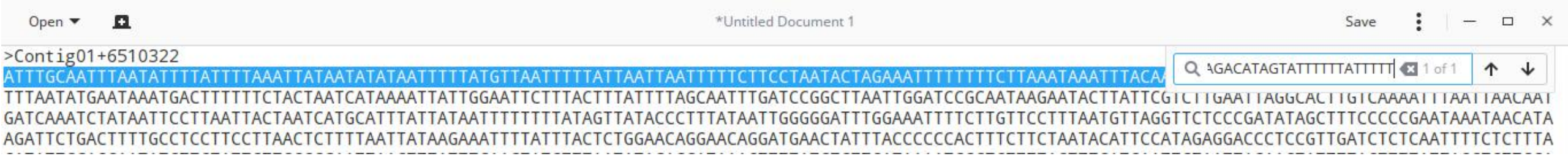
Pressione “Ctrl+F” e um campo de busca vai aparecer no canto superior direito da tela.

Vamos selecionar a primeira linha do primeiro contig.

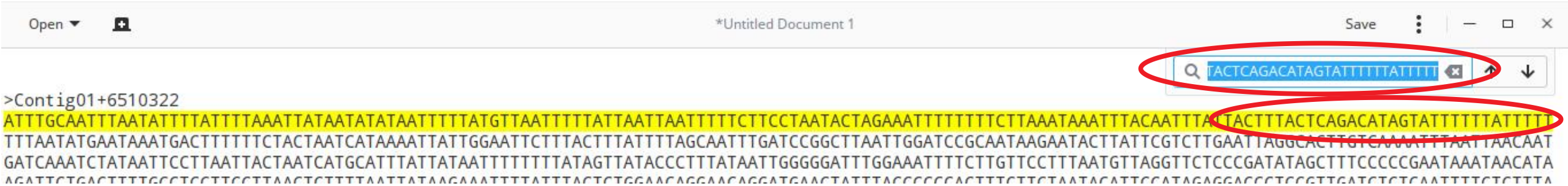
Open *Untitled Document 1 Save

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
```

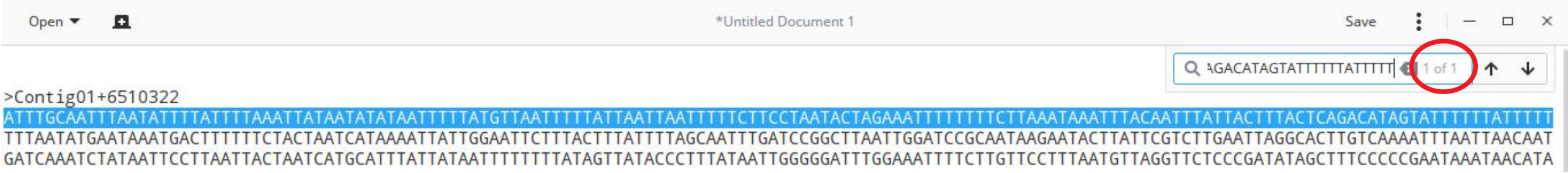
Aperte “Ctrl+F” e a sequência já estará colada na caixa de busca que irá aparecer.



O final da sequência selecionada vai ficar visível na caixa de busca.



Para ter certeza que nenhuma subsequência dessa sequência maior esteja em outro contig, temos que ir deletando ela usando “backspace” e ver se ela aparece em outros contigs. Vemos isso por meio do indicador “1 of 1” à direita da caixa de busca.



Removendo nucleotídeos… até agora não encontramos outras regiões parecidas.

Open *Untitled Document 1 Save

AATACTAGAAATTTTTTTTCTT 1 of 1

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
```

Quase todos os nucleotídeos removidos e ainda nada.

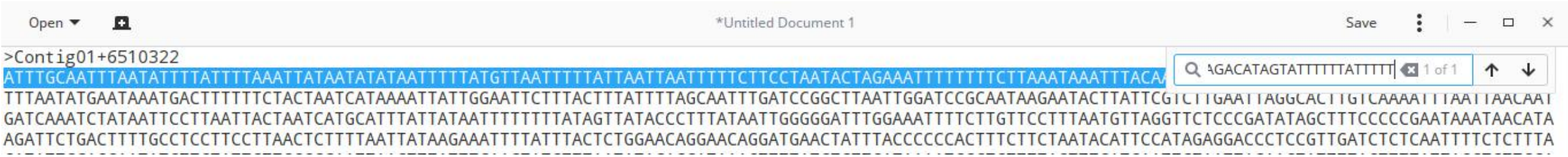
Open *Untitled Document 1 Save

CAATTTAATATTTTATTTTAAA 1 of 1

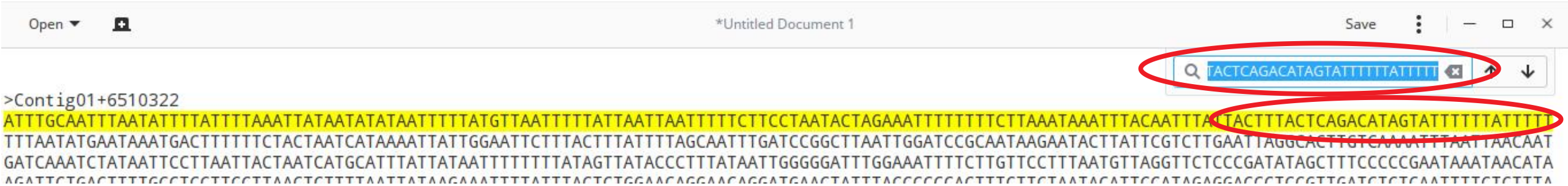
```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
```

Agora apareceram várias regiões iguais (193), mas apenas porque temos poucos nucleotídeos. Essa sobreposição não é boa o bastante para unirmos dois contigs diferentes.

Para termos certeza absoluta que a região do começo não sobrepõe com nenhuma outra, devemos fazer um segundo passo. Selecione novamente a linha inteira e pressione “Ctrl+F”

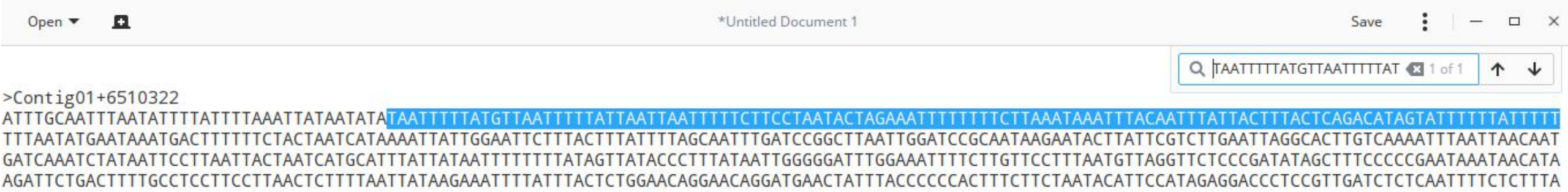


O final da sequência selecionada vai ficar visível na caixa de busca.



Na caixa de busca, aperte “Ctrl+ ⌘ ” ou simplesmente “⌘ ”várias vezes, até chegar ao começo da sequência selecionada

Daí, comece a deletar a sequência na caixa do “Ctrl+F” pressionando a tecla “Del” ou “Delete”, dependendo do do seu teclado.



Mais uma vez, estamos chegando ao final da seleção e nada.

Open *Untitled Document 1 Save

ACTTTACTCAGACATAGTATT 1 of 1

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
```

Com isso, confirmamos que não há região de sobreposição do começo desse contig com os outros contigs.

Open *Untitled Document 1 Save

TTTTT 1 of 66

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTATTATATTTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTGTTTCTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
```

Vamos repetir o processo com a última linha do primeiro contig. Primeiro, vamos usar o “backspace”.

Obs: se a ultima linha tiver poucos nucleotídeos, pode usar a penúltima.

Open *Untitled Document 1 Save

ATTATACTATTATTTATTTTATCTTTT

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
```

Deletando nucleotídeos com o “backspace”...

Open ▾ *Untitled Document 1 Save

ATTCTTTTATTCCTCAAATAAT 1 of 1

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
```

…e não encontramos sobreposição novamente.

Open ▾ *Untitled Document 1 Save

TTAA 1 of 142

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
```

Por último, vamos retirar os nucleotídeos usando o “del”

Open ▾ *Untitled Document 1 Save

TTAAATTGTCAATTTAAAATTATTTAT

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
```

Removendo nucleotídeos com o “del”…

Open ▾ *Untitled Document 1 Save

TCTTTTATTCCTCAAATAATAC 1 of 1

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
```

Até que finalmente encontramos uma sobreposição válida entre dois contigs.

Open *Untitled Document 1 Save

CTATTTATAATAATTATTTTAA 1 of 2

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAAATTGAAAATGATACTAAATATAT
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
```

Daí, basta excluirmos a região de sobreposição...

Open *Untitled Document 1 Save

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTT
>Contig01+6510322
CTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAACATATTTAAATTTTATAAAAATAAAATTTTATACTTTCATTGAA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCT
AATTTAAATCTTTAATTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAG
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACT
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTAT
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGC
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAA
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTGTTTCTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAA
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATC
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTC
```

Undo
Redo
Cut
Copy
Paste
Delete
Select All
Change Case

E unirmos os dois contigs.

Open *Untitled Document 1 Save

```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTA[illegible]AAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTA[illegible]TAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGG[illegible]CATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGA[illegible]TTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTA[illegible]AAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTT[illegible]AATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTCCATTTCAACCATCTTCATCATACTCACATTTCTACATCTACTATCACTATTCTTTTTATTTCAATATATTCATCAACATTTTAATTTTATTTATTATATACTATAATCTTAT
```

Nesse exemplo, ainda sobrou um pequeno contig. Vamos pegar a primeira linha dele e ir retirando nucleotídeos usando o “backspace”

TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTC(
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAG(
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
AGAATAAAATTTTATTTCTTCCTATACACTTATATCTTCTTCCCCTCTCTATTAACTATCACCCCTATTACTCACTTTTAAATACCACCCATCAAAATTTAATTATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
ATTGTTAATAATATAAATGAAATATTTTCTAGTTGAAGTAATTTTTAATGAACTTCTAATTCATTTTTTTATGATTTTTTTCATATTACTTCTTTCTTACTTTTAGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATAAT
AAAAATAGTTTACAAAAACTTATTTAAAAATTTTAATTAATTTTCTTAATATCTTCAATATTATACTTTATATATAAGCTATTTAAATTAATAACATATATTTACAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTT
AAAATTAATTATTTTATATTTACT

>Contig01+6510322

TATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACTAAGTATAAAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATCTCTCTTCTTAATTTGTTAAAAGGATGATAAATTCATG
TATATAATATACTTAAAAATCATATAGATCTCACAAAATACCCTAAATAATATAATTTTAAATATTTTTTAATATTAATCACTATAATAATTCCCACAATTCCCCCCAATAAATATATACCCAAAGTAATTAATTTAACTTTTATAGATA

Encontramos uma região de sobreposição. Entretanto, ela não vai até o final do nosso contig por haver uma **quebra de linha.**

TATTTATAAATATATAATACTT 2 of 2

TTTAATATGAATAAATGACTTTTTTCTACTAATCAAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCC
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGC
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATATTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTTATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGATTTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCCCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTATACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTCATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATACATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
AGAATAAAATTTTATTTCTTCCTATACACTTATATCTTCTTCCCCTCTCTATTAACTATCACCCCTATTACTCACTTTTAAATACCACCCATCAAAATTTAATTATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
ATTGTTAATAATATAAATGAAATATTTTCTAGTTGAAGTAATTTTTAATGAACTTCTAATTCATTTTTTTATGATTTTTTTCATATTACTTCTTTCTTACTTTTAGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATTT
AAAAATAGTTTACAAAAACTTATTTAAAAATTTTAATTAATTTTCTTAATATCTTCAATATTATACTTTATATATAAGCTATTTAAATTAATAACATATATTTACAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTT
AAAATTAATTATTTTATATTTACT
>Contig01+6510322
TATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACTAAGTATAAAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATCTCTCTTCTTAATTTGTTAAAAGGATGATAAATTCATG

Para contornar isso, precisamos ir até **antes do primeiro nucleotídeo da linha seguinte** e pressionar “backspace”

Daí, conseguimos retirar a quebra de linha...

```
TTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
GGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
AACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
CATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
AAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
CGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
ATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
GATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
ATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
TTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
CCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
AACTATCACCCCTATTACTCACTTTTAAATACCACCCATCAAAATTTAATTATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
TTCTAATTCATTTTTTTATGATTTTTTTCATATTACTTCTTTCTTACTTTTAGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATAAT
TTCAATATTATACTTTATATATAAGCTATTTAAATTAATAACATATATTTACAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACT

AAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATCTCTCTTCTTAATTTGTTAAAAGGATGATAAATTCATG
AATTTTAAATATTTTTTAATATTAATCACTATAATAATTCCCACAATTCCCCCCAATAAATATATACCCAAAGTAATTAATTTAACTTTTATAGATA
ATTACTCTAAAAACTAATAAAATTATTATAGAAAAATTTATTAATTTCTCTTC
```

Plain Text ▾ Tab Width: 8 ▾ Ln 18, Col 126 ▾ INS

..e encontrar a verdadeira região de sobreposição, que é um pouco maior.

AATTAATTATTTTATATTTACT 2 of 2

ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAA
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTC
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGG
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTAA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
AGAATAAAATTTTATTTCTTCCTATACACTTATATCTTCTTCCCCTCTCTATTAACTATCACCCCTATTACTCACTTTTAAATACCACCCATCAAAATTTAATTATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
ATTGTTAATAATATAAATGAAATATTTTCTAGTTGAAGTAATTTTTAATGAACTTCTAATTCATTTTTTTATGATTTTTTTCATATTACTTCTTTCTTACTTTTAGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATAAT
AAAAATAGTTTACAAAAACTTATTTAAAAATTTTAATTAATTTTCTTAATATCTTCAATATTATACTTTATATATAAGCTATTTAAATTAATAACATATATTTACAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTTAA
>Contig01+6510322
TATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACTAAGTATAAAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATCTCTCTTCTTAATTTGTTAAAAGGATGATAAATTCATG

Daí, mais uma vez, devemos deletar a região de sobreposição...

```
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAA                                                CTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTA                                                GATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTA                                                CCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTT                                                AAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTAC                                                ATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATT                                                ATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATT                                                TTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATT                                                TGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGAT                                                CATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTT                                                TTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
AGAATAAAATTTTATTTCTTCCTATACACTTATATCTTCTTCCCCTCTCTATTAACTATCACCCCTATTACTCACTTTTAAATACCA                                                ATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
ATTGTTAATAATATAAATGAAATATTTTCTAGTTGAAGTAATTTTTAATGAACTTCTAATTCATTTTTTTATGATTTTTTTCATATT                                                AGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATAAT
AAAAATAGTTTACAAAAACTTATTTAAAAATTTTAATTAATTTTCTTAATATCTTCAATATTATACTTTATATATAAGCTATTTAAA                                                CAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTTAA
>Contig01+6510322
TATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACTAAGTATAAAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATCTCTCTTCTTAATTTGTTAAAAGGATGATAAATTCATG
TATATAATATACTTAAAAATCATATAGATCTCACAAAATACCCTAAATAATATAATTTTAAATATTTTTTAATATTAATCACTATAATAATTCCCACAATTCCCCCCAATAAATATATACCCAAAGTAATTAATTTAACTTTTATAGATA
AATAAACTATATAAAAATCAAAAAAAAATACTCAATTTATTAAAGCTCCTATTATTACTCTAAAAACTAATAAAATTATTATAGAAAAATTTATTAATTTCTCTTC
```

Undo
Redo
Cut
Copy
Paste
Delete
Select All
Change Case

Plain Text Tab Width: 8 Ln 38, Col 47 INS

...e **unir os contigs**, gerando uma sequência única.

CTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
TTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
TCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
TTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
GCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
TTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
CCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
CTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
CATCAAAATTTAATTATTATTAAATAAATTAATATTAACTAAAACCAAAAAGAGGTATATT
TTCTTTCTTACTTTTAGTAACTTAAATATAGTTTATTTTTAAAACATTATTTTTTCATAAT
AATAACATATATTTACAAATATAATTATTAACATAAATGTATTTATAAATATATAATACTTAAAATTAATTATTTTATATTTACTAAGTATAAAATTTATTAAATAATATTTTTCTAACTTTTCAATTCAACTCTTATCTACCTTATCAATC
ATTCCCACAATTCCCCCCAATAAATATATACCCAAAGTAATTAATTTAACTTTTATAGATA
TTTATTAATTTCTCTTC

Ponto no qual unimos os dois contigs

Em suma, temos que repetir esse mesmo procedimento de encontrar sobreposições para todos os contigs gerados. Assim, podemos estender ainda mais a montagem do NOVOPlasty, o que geralmente torna a próxima etapa mais rápida.

Caso não dê para fundir todos os contigs, manter somente o maior e deletar o(s) menor(es).

Por último, devemos voltar ao início do arquivo fasta e mudar a primeira linha dele para “>Nome_da_espécie-PARTIAL “

Open *Untitled Document 1 Save

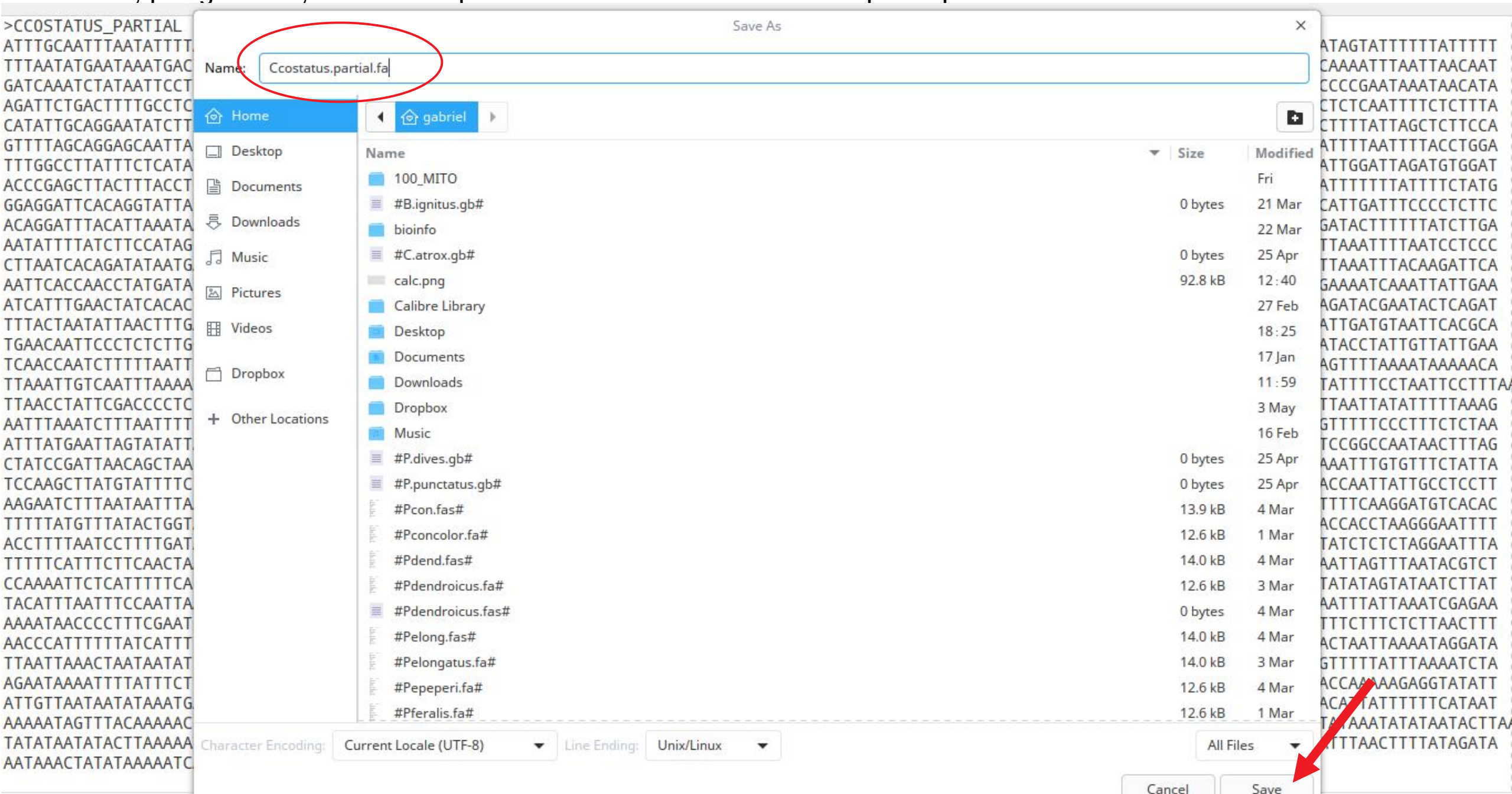
```
>Contig01+6510322
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
AACCCATTTTTTATCATTTTATTATTAATTTTTTTAAGTAGTCTTATTTTTAGCCTTTATTTAGAGTATAACGAAAATTCAATTGATTGAAAAATTTAATTTTACATATTTATATATACACATATTAAATTTACTAATTAAAATAGGATA
TTAATTAAACTAATAATATTTAAATTGCATTTAAAAATAATAACCTTAAATTATATATCTTACTGGCATTACACATCTATAATCTTTATTCTTAAAGTAATGATTTACACTTTAATTTCGAATTAAATATTAGTTTTTATTTAAAATCTA
```

Por último, devemos voltar ao início do arquivo fasta e mudar a primeira linha dele para “>Nome_da_espécie-PARTIAL “

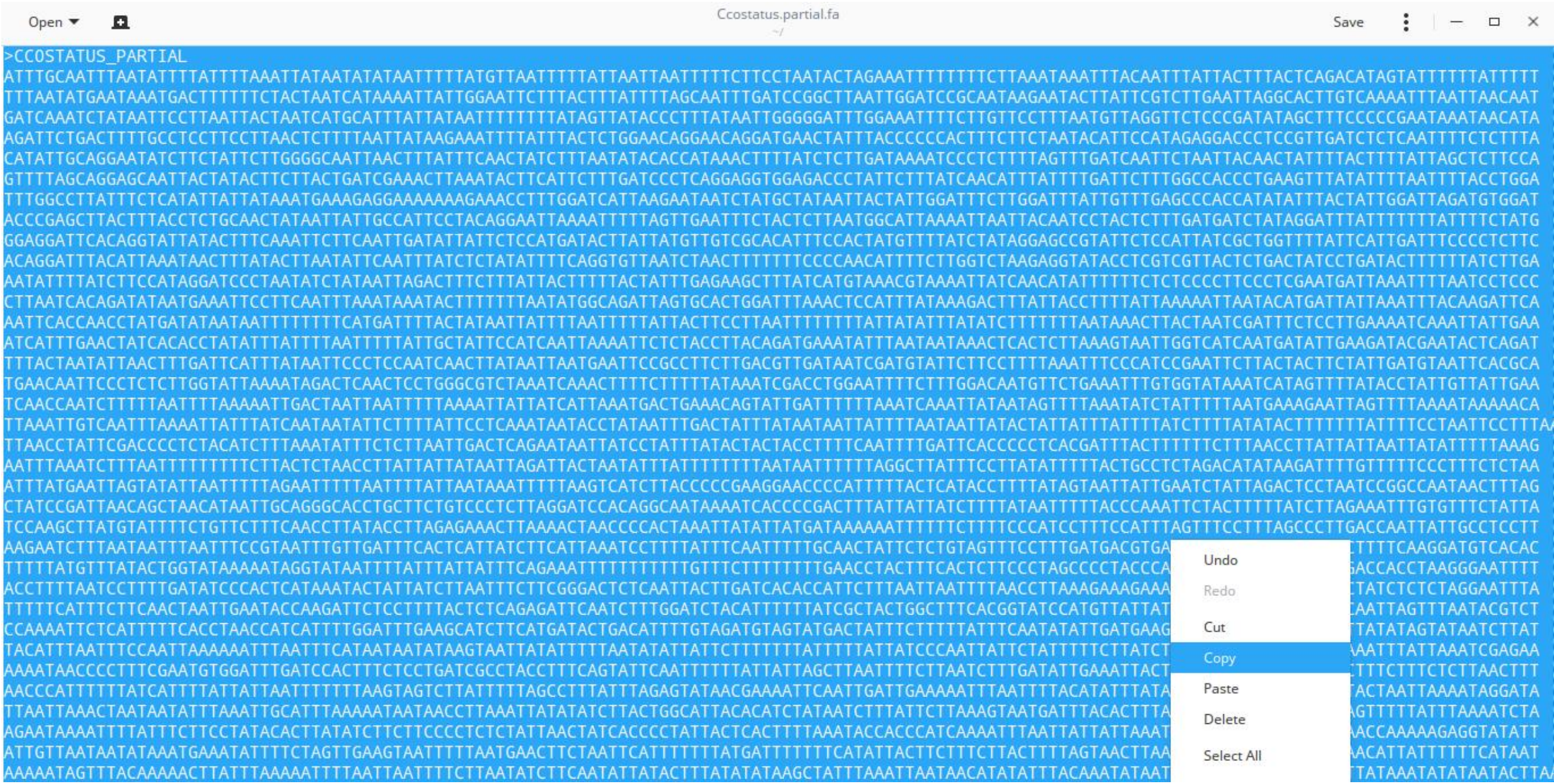
Open *Untitled Document 1 Save

```
>CCOSTATUS_PARTIAL
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTTTCCTAATTCCTTTA
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
ACCTTTTAATCCTTTTGATATCCCACTCATAAATACTATTATCTTAATTTCTTCGGGACTCTCAATTACTTGATCACACCATTCTTTAATTAATTTTAACCTTAAAGAAAGAAAAAATAGATTAATCTTAACTATCTCTCTAGGAATTTA
TTTTTCATTTCTTCAACTAATTGAATACCAAGATTCTCCTTTTACTCTCAGAGATTCAATCTTTGGATCTACATTTTTTATCGCTACTGGCTTTCACGGTATCCATGTTATTATTGGATCGATTTTTCTTTCAATTAGTTTAATACGTCT
CCAAAATTCTCATTTTTCACCTAACCATCATTTTGGATTTGAAGCATCTTCATGATACTGACATTTTGTAGATGTAGTATGACTATTTCTTTTTATTTCAATATATTGATGAAGATTTTAATTTTATTTATTTATATAGTATAATCTTAT
TACATTTAATTTCCAATTAAAAAATTTAATTTCATAATAATATAAGTAATTATATTTTTAATATATTATTCTTTTTTTATTTTTATTATCCCAATTATTCTATTTTTCTTATCTCTAATTATTTCTAAAAAAAATTTATTAAATCGAGAA
AAAATAACCCCTTTCGAATGTGGATTTGATCCACTTTCTCCTGATCGCCTACCTTTCAGTATTCAATTTTTTATTATTAGCTTAATTTTCTTAATCTTTGATATTGAAATTACTCTACTAATTCCTCTAATCTTTCTTTCTCTTAACTTT
```

E, por garantia, salvar o arquivo com o nome “Nome da espécie.partial.fa”



Agora, copie toda a sequencia desse novo arquivo (Ctrl+a para selecionar tudo e Ctrl+c para copiar)



Voltando ao terminal, vamos criar um arquivo fasta na pasta da montagem NOVOPlasty usando o nano. Cole a sequencia (Ctrl+Shift+C) e salve.

nano nome_da_espécie.partial.fa

gabriel@Leveza: ~/100_MITO/CCOSTATUS-NOVOPLASTY

```
gabriel@Leveza:~/100_MITO/CCOSTATUS-NOVOPLASTY$ nano Ccostatus.partial.fa
```

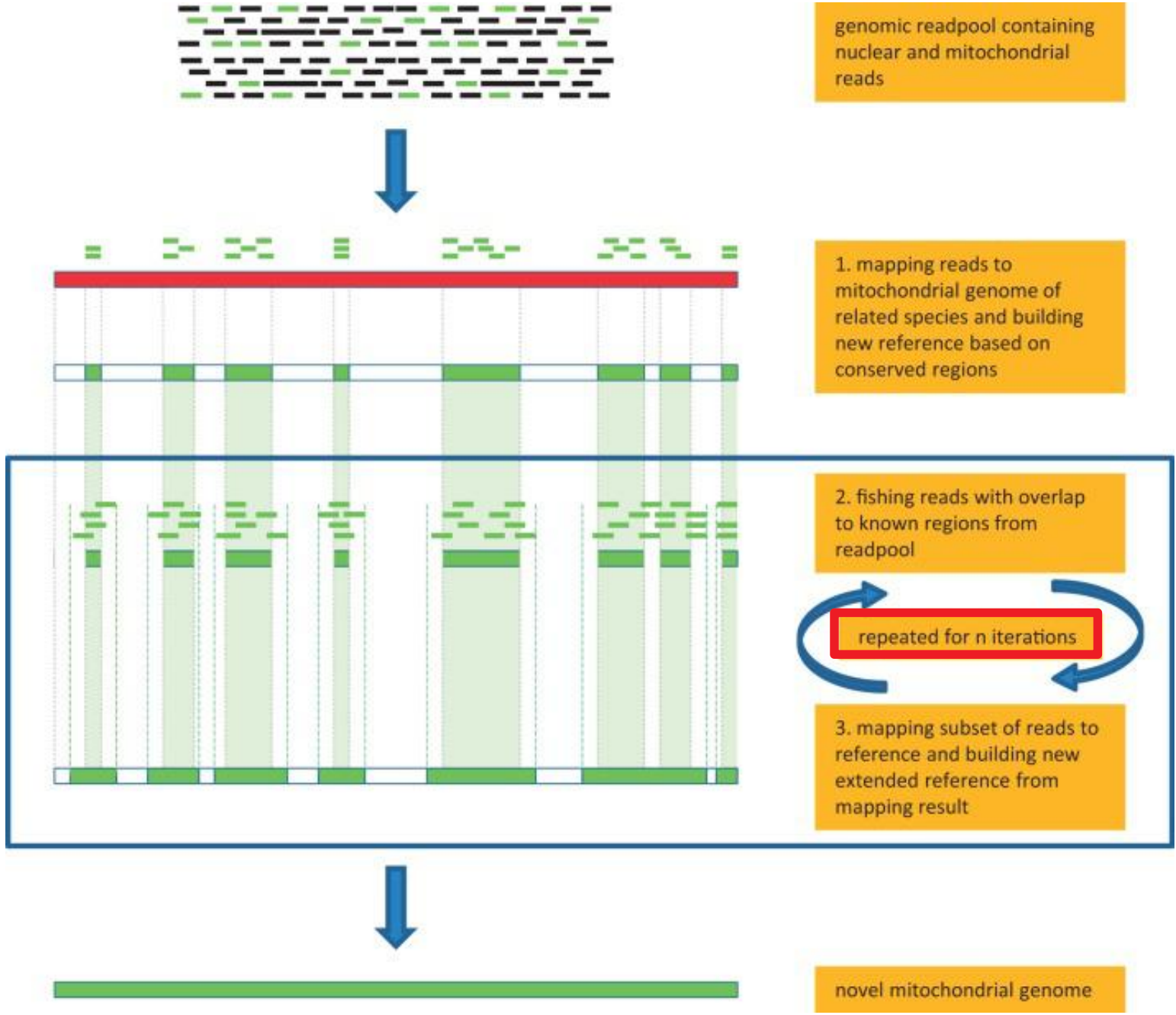
gabriel@Leveza: ~/100_MITO/CCOSTATUS-NOVOPLASTY

```
  GNU nano 2.8.6                    File: Ccostatus.partial.fa                    Modified

>CCOSTATUS_PARTIAL
ATTTGCAATTTAATATTTTATTTTAAATTATAATATATAATTTTTATGTTAATTTTTATTAATTAATTTTTCTTCCTAATACTAGAAATTTTTTTTCTTAAATAAATTTACAATTTATTACTTTACTCAGACATAGTATTTTTTATTTTT
TTTAATATGAATAAATGACTTTTTTCTACTAATCATAAAATTATTGGAATTCTTTACTTTATTTTAGCAATTTGATCCGGCTTAATTGGATCCGCAATAAGAATACTTATTCGTCTTGAATTAGGCACTTGTCAAAATTTAATTAACAAT
GATCAAATCTATAATTCCTTAATTACTAATCATGCATTTATTATAATTTTTTTTATAGTTATACCCTTTATAATTGGGGGATTTGGAAATTTTCTTGTTCCTTTAATGTTAGGTTCTCCCGATATAGCTTTCCCCCGAATAAATAACATA
AGATTCTGACTTTTGCCTCCTTCCTTAACTCTTTTAATTATAAGAAATTTTATTTACTCTGGAACAGGAACAGGATGAACTATTTACCCCCCACTTTCTTCTAATACATTCCATAGAGGACCCTCCGTTGATCTCTCAATTTTCTCTTTA
CATATTGCAGGAATATCTTCTATTCTTGGGGCAATTAACTTTATTTCAACTATCTTTAATATACACCATAAACTTTTATCTCTTGATAAAATCCCTCTTTTAGTTTGATCAATTCTAATTACAACTATTTTACTTTTATTAGCTCTTCCA
GTTTTAGCAGGAGCAATTACTATACTTCTTACTGATCGAAACTTAAATACTTCATTCTTTGATCCCTCAGGAGGTGGAGACCCTATTCTTTATCAACATTTATTTTGATTCTTTGGCCACCCTGAAGTTTATATTTTAATTTTACCTGGA
TTTGGCCTTATTTCTCATATTATTATAAATGAAAGAGGAAAAAAAGAAACCTTTGGATCATTAAGAATAATCTATGCTATAATTACTATTGGATTTCTTGGATTTATTGTTTGAGCCCACCATATATTTACTATTGGATTAGATGTGGAT
ACCCGAGCTTACTTTACCTCTGCAACTATAATTATTGCCATTCCTACAGGAATTAAAATTTTTAGTTGAATTTCTACTCTTAATGGCATTAAAATTAATTACAATCCTACTCTTTGATGATCTATAGGATTTATTTTTTTATTTTCTATG
GGAGGATTCACAGGTATTATACTTTCAAATTCTTCAATTGATATTATTCTCCATGATACTTATTATGTTGTCGCACATTTCCACTATGTTTTATCTATAGGAGCCGTATTCTCCATTATCGCTGGTTTTATTCATTGATTTCCCCTCTTC
ACAGGATTTACATTAAATAACTTTATACTTAATATTCAATTTATCTCTATATTTTCAGGTGTTAATCTAACTTTTTTTCCCCAACATTTTCTTGGTCTAAGAGGTATACCTCGTCGTTACTCTGACTATCCTGATACTTTTTTATCTTGA
AATATTTTATCTTCCATAGGATCCCTAATATCTATAATTAGACTTTCTTTATTACTTTTTACTATTTGAGAAGCTTTATCATGTAAACGTAAAATTATCAACATATTTTTTCTCTCCCCTTCCCTCGAATGATTAAATTTTAATCCTCCC
CTTAATCACAGATATAATGAAATTCCTTCAATTTAAATAAATACTTTTTTTAATATGGCAGATTAGTGCACTGGATTTAAACTCCATTTATAAAGACTTTATTACCTTTTATTAAAAATTAATACATGATTATTAAATTTACAAGATTCA
AATTCACCAACCTATGATATAATAATTTTTTTTCATGATTTTACTATAATTATTTTAATTTTTATTACTTCCTTAATTTTTTTTATTATATTTATATCTTTTTTTAATAAACTTACTAATCGATTTCTCCTTGAAAATCAAATTATTGAA
ATCATTTGAACTATCACACCTATATTTATTTTAATTTTTATTGCTATTCCATCAATTAAAATTCTCTACCTTACAGATGAAATATTTAATAATAAACTCACTCTTAAAGTAATTGGTCATCAATGATATTGAAGATACGAATACTCAGAT
TTTACTAATATTAACTTTGATTCATTTATAATTCCCTCCAATCAACTTATAATTAATGAATTCCGCCTTCTTGACGTTGATAATCGATGTATTCTTCCTTTTAAATTTCCCATCCGAATTCTTACTACTTCTATTGATGTAATTCACGCA
TGAACAATTCCCTCTCTTGGTATTAAAATAGACTCAACTCCTGGGCGTCTAAATCAAACTTTTCTTTTTATAAATCGACCTGGAATTTTCTTTGGACAATGTTCTGAAATTTGTGGTATAAATCATAGTTTTATACCTATTGTTATTGAA
TCAACCAATCTTTTTAATTTTAAAAATTGACTAATTAATTTTTAAAATTATTATCATTAAATGACTGAAACAGTATTGATTTTTTAAATCAAATTATAATAGTTTTAAATATCTATTTTTAATGAAAGAATTAGTTTTAAAATAAAAACA
TTAAATTGTCAATTTAAAATTATTTATCAATAATATTCTTTTATTCCTCAAATAATACCTATAATTTGACTATTTATAATAATTATTTTAATAATTATACTATTTATAATAATTATTTTAATAATTATACTATTATTTATTTTATCTTTTATATACTTTTTTTATTT$
TTAACCTATTCGACCCCTCTACATCTTTAAATATTTCTCTTAATTGACTCAGAATAATTATCCTATTTATACTACTACCTTTTCAATTTTGATTCACCCCCTCACGATTTACTTTTTTCTTTAACCTTATTATTAATTATATTTTTAAAG
AATTTAAATCTTTAATTTTTTTTTCTTACTCTAACCTTATTATTATAATTAGATTACTAATATTTATTTTTTTTAATAATTTTTTAGGCTTATTTCCTTATATTTTTACTGCCTCTAGACATATAAGATTTTGTTTTTCCCTTTCTCTAA
ATTTATGAATTAGTATATTAATTTTTAGAATTTTTAATTTTATTAATAAATTTTTAAGTCATCTTACCCCCGAAGGAACCCCATTTTTACTCATACCTTTTATAGTAATTATTGAATCTATTAGACTCCTAATCCGGCCAATAACTTTAG
CTATCCGATTAACAGCTAACATAATTGCAGGGCACCTGCTTCTGTCCCTCTTAGGATCCACAGGCAATAAAATCACCCCGACTTTATTATTATCTTTTATAATTTTTACCCAAATTCTACTTTTTATCTTAGAAATTTGTGTTTCTATTA
TCCAAGCTTATGTATTTTCTGTTCTTTCAACCTTATACCTTAGAGAAACTTAAAACTAACCCCACTAAATTATATTATGATAAAAAATTTTTTCTTTTCCCATCCTTTCCATTTAGTTTCCTTTAGCCCTTGACCAATTATTGCCTCCTT
AAGAATCTTTAATAATTTAATTTCCGTAATTTGTTGATTTCACTCATTATCTTCATTAAATCCTTTTATTTCAATTTTTGCAACTATTCTCTGTAGTTTCCTTTGATGACGTGATGTAGCTCGAGAATCTACTTTTCAAGGATGTCACAC
TTTTTATGTTTATACTGGTATAAAAATAGGTATAATTTTATTTATTATTTCAGAAATTTTTTTTTTTGTTTCTTTTTTTTGAACCTACTTTCACTCTTCCCTAGCCCCTACCCAAGAAATTGGTCAATCTTGACCACCTAAGGGAATTTT
```

Agora, independentemente de qual foi o resultado da sua montagem pelo NOVOPlasty, você tem uma sequência mitocondrial em formato fasta que será usada como referência para uma segunda montagem usando o programa **MITObim**.

Esse programa é muito usado para fechar gaps e para isso ele faz várias montagens consecutivas (**iterações**).



O MITObim não funciona com dois arquivos fastq simultaneamente e também não reconhece o cabeçalho das reads no padrão SRA. Logo, precisamos usar o fastq-dump novamente para gerar um fastq único com cabeçalho modificado.

Além disso, por limitações de tempo e recursos computacionais, iremos usar um número limitado de reads para essa última montagem. A princípio, podemos usar 10 milhões de reads.

Para fazer isso, digite:

/caminho/para/fastq-dump -X 5000000 --split-spot --defline-seq '@$ac-$sn/$ri' --defline-qual '+' -O ./ **/caminho/para/arquivo.sra** &

X 5000000 : Imprime os primeiros 5 milhões de spots, que geralmente significam 10milhões de reads para dados paired end. Esse valor pode ser mudado para obter o número de reads que você quiser.

Split-spot : coloca todos os dados em um único arquivo.

defline-seq e defline-qual : mudam o cabeçalho

Ao final da conversão, usando “ls -ltrh”, podemos ver que apenas um arquivo foi gerado. Ele será usado pelo MITObim para a segunda montagem. O

```
-rw-rw-r-- 1 gabriel gabriel  939 mai  3 17:51 Wrhopalogaster.seed.fa
-rw-rw-r-- 1 gabriel gabriel  58K mai  3 17:53 SRR1145745.Saequinoctialis.fastq
-rw-rw-r-- 1 gabriel gabriel  76K mai  3 17:54 SRR4342885.Tlaticeps.fastq
-rw-rw-r-- 1 gabriel gabriel  78K mai  3 17:54 SRR5907418.Wrhopalogaster.fastq
drwxr-xr-x 2 gabriel gabriel 4,0K mai  3 18:07 NOVOPLASTY_Saequinoctalis
drwxr-xr-x 2 gabriel gabriel 4,0K mai  3 18:44 NOVOPLASTY_Wrhopalogaster
drwxr-xr-x 2 gabriel gabriel 4,0K mai  3 18:53 NOVOPLASTY_Tlaticeps
-rw-rw-r-- 1 gabriel gabriel  10M mai 15 12:18 SRR3168941.fastq
-rw-rw-r-- 1 gabriel gabriel 267M mai 15 14:11 SRR3168941.Ccostatus_2.fastq
-rw-rw-r-- 1 gabriel gabriel 267M mai 15 14:11 SRR3168941.Ccostatus_1.fastq
-rw-r--r-- 1 gabriel gabriel 339M mai 15 19:55 SRR3168940.Ccostatus.200bp.sra
-rw------- 1 gabriel gabriel 1,4K mai 16 00:07 nohup.out
drwxr-xr-x 2 gabriel gabriel 4,0K mai 16 15:00 NOVOPLASTY_Salternans
drwxrwxr-x 2 gabriel gabriel 4,0K mai 16 17:12 DFERRUGINOSA-NOVOPLASTY
drwxrwxr-x 2 gabriel gabriel 4,0K mai 19 13:29 CCOSTATUS-NOVOPLASTY
-rw-rw-r-- 1 gabriel gabriel 773M jun  2 11:15 SRR5651490.Salternans.sra
-rw-rw-r-- 1 gabriel gabriel 3,2G jun  2 12:03 SRR5651490.Salternans.fastq
gabriel@Leveza:~/100_MITO$
```

Agora podemos finalmente usar o MITObim. O programa está em **/bioinfo/MITObim-master/MITObim.pl** em nossos servidores. Pode ser baixado em **https://github.com/chrishah/MITObim**

Ele depende de um outro programa, o MIRA, que está em **/bioinfo/mira_4.0.2_linux-gnu_x86_64_static/bin/ .** No momento MITObim só é compatível com a versão 4 do mira, que pode ser baixada em **https://sourceforge.net/projects/mira-assembler/files/MIRA/stable/**

Para rodar o MITObim, você terá que fornecer tanto o caminho do próprio MITObim quanto do MIRA (se ele não estiver instalado na variável PATH). De preferência, rode ele de dentro da pasta da montagem NOVOPlasty.

Digite:

nohup **/caminho/para/MITObim.pl** --mirapath **/caminho/para/mira4.0.2** -end 100 -quick **/caminho/para/sequencia_mitocondrial.fa** -sample **nome_da_espécie** -ref reference -readpool **/caminho/para/arquivo.fastq** --pair --clean >mitobim.out 2>mitobim.err &

Ex: nohup /bioinfo/MITObim-master/MITObim.pl -end 100 --mirapath /bioinfo/mira_4.0.2_linux-gnu_x86_64_static/bin/ -quick /home/gabriel/100_MITO/NOVOPLASTY_Salternans/Salternans.mito.circ.fa -sample Salternans -ref reference -readpool /home/gabriel/100_MITO/SRR5651490.Salternans.fastq --pair --clean >mitobim.out 2>mitobim.err &

Usando o comando “top”, podemos ver que o MITObim está rodando (mirabait). Agora é só esperar ele terminar.

Se houver algum erro, abra os arquivos “mitobim.out” ou “mitobim.err” e veja qual o problema. Geralmente, é algo simples de se resolver, como algum caminho que não foi digitado corretamente, ou alguma pasta que precise ser removida antes de rodar o programa novamente (geralmente é a pasta **iteration0**. Para remover, digite **“rm -r iteration0”**.

```
gabriel@Leveza: SRA    gabriel@Leveza: NOVOPLASTY_Salternans
Tasks: 353 total,   2 running, 351 sleeping,   0 stopped,   0 zombie
%Cpu(s):  0,0 us,  0,0 sy,  0,0 ni, 99,9 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
KiB Mem : 32917216 total, 19197252 free,  2473532 used, 11246432 buff/cache
KiB Swap:  2097148 total,  2097148 free,        0 used. 29937904 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
14562 gabriel   20   0  301168 270272   4860 R 100,0  0,8   0:23.55 mirabait
14564 gabriel   20   0   46604   4064   3196 R   6,2  0,0   0:00.01 top
    1 root      20   0  220432   8684   6276 S   0,0  0,0   0:07.72 systemd
    2 root      20   0       0      0      0 S   0,0  0,0   0:00.01 kthreadd
    4 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 kworker/0:0H
    6 root       0 -20       0      0      0 S   0,0  0,0   0:00.00 mm_percpu_wq
    7 root      20   0       0      0      0 S   0,0  0,0   0:00.01 ksoftirqd/0
    8 root      20   0       0      0      0 S   0,0  0,0   1:16.72 rcu_sched
    9 root      20   0       0      0      0 S   0,0  0,0   0:00.00 rcu_bh
   10 root      rt   0       0      0      0 S   0,0  0,0   0:00.00 migration/0
   11 root      rt   0       0      0      0 S   0,0  0,0   0:00.54 watchdog/0
   12 root      20   0       0      0      0 S   0,0  0,0   0:00.00 cpuhp/0
   13 root      20   0       0      0      0 S   0,0  0,0   0:00.00 cpuhp/1
   14 root      rt   0       0      0      0 S   0,0  0,0   0:00.57 watchdog/1
   15 root      rt   0       0      0      0 S   0,0  0,0   0:00.01 migration/1
   16 root      20   0       0      0      0 S   0,0  0,0   0:00.21 ksoftirqd/1
```

Cheque se a montagem terminou usando o “top” ou abrindo as últimas linhas do arquivo “mitobim.out”

Para fazer isso, digite “tail -n20 mitobim.out”, e as ultimas 20 linhas do arquivo seerão impressas na tela

Se a mensagem “Final assembly result will be written to file” aparecer, a montagem terminou.

```
gabriel@Leveza: NOVOPLASTY_Salternans
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans$ tail -n20 mitobim.out
re-running mapping assembly using MIRA

readpool contains 67551 reads
assembly contains 1 contig(s)
contig length: 17493

now removing directory iteration0

MITObim has reached a stationary read number after 2 iterations!!

Final assembly result will be written to file: /home/gabriel/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans_reference-it2_noIUPAC.fasta

if you found MITObim useful, please cite:
Hahn C, Bachmann L and Chevreux B. (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads -
a baiting and iterative mapping approach. Nucl. Acids Res. 41(13):e129. doi: 10.1093/nar/gkt371

jun  2 13:05:50
```

Vamos ver o que temos na pasta da montagem agora. Use “ls -ltrh”.

O MITObim gerou duas pastas aqui, “iteration1” e “iteration2”. Essa foi uma montagem que terminou bem rápido, mas em vários casos o programa vai fazer várias iterações antes de finalizar a montagem.

Nós sempre iremos trabalhar com a última iteração gerada pelo programa. No caso, a iteration2.

Vá para a pasta iteration2 usando “cd iteration2”

```
gabriel@Leveza: NOVOPLASTY_Salternans
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans$ ls -lrht
total 13M
-rw-rw-r-- 1 gabriel gabriel  17K mai  3 18:04 Circularized_assembly_1_SALTERANS.fasta
-rw-r--r-- 1 gabriel gabriel 4,8K mai  3 18:14 config.txt
-rw-rw-r-- 1 gabriel gabriel   22 mai  3 18:14 NOVOP.err
-rw-rw-r-- 1 gabriel gabriel  18K mai  3 18:52 Option_2_SALTERANS.fasta
-rw-rw-r-- 1 gabriel gabriel  18K mai  3 18:52 Option_1_SALTERANS.fasta
-rw-rw-r-- 1 gabriel gabriel  83K mai  3 18:52 NOVOP.out
-rw-rw-r-- 1 gabriel gabriel  36K mai  3 18:52 Merged_contigs_SALTERANS.txt
-rw-rw-r-- 1 gabriel gabriel 2,1K mai  3 18:52 log_SALTERANS.txt
-rw-rw-r-- 1 gabriel gabriel 2,5K mai  3 18:52 contigs_tmp_SALTERANS.txt
-rw-rw-r-- 1 gabriel gabriel  19K mai  3 18:52 Contigs_1_SALTERANS.fasta
-rw-rw-r-- 1 gabriel gabriel 6,3M mai  3 18:52 Assembled_reads_SALTERANS_R2.fasta
-rw-rw-r-- 1 gabriel gabriel 6,3M mai  3 18:52 Assembled_reads_SALTERANS_R1.fasta
-rw-rw-r-- 1 gabriel gabriel  17K mai 16 14:46 Salternans.mito.circ.fa
-rw------- 1 gabriel gabriel 2,8K jun  2 12:35 nohup.out
-rw-rw-r-- 1 gabriel gabriel   22 jun  2 12:52 mitobim.err
drwxrwxr-x 4 gabriel gabriel 4,0K jun  2 12:59 iteration1
drwxrwxr-x 4 gabriel gabriel 4,0K jun  2 13:05 iteration2
-rw-rw-r-- 1 gabriel gabriel 3,0K jun  2 13:05 mitobim.out
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans$
```

Essa pasta tem vários arquivos, incluindo a montagem final do mitobim.

O nome dela será algo parecido com **nome_da_espécie_reference-itX_noIUPAC.fasta**

```
gabriel@Leveza: iteration2
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$ ls -ltrh
total 46M
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:01 backbone_it2_initial_reference.fna
-rw-rw-r-- 1 gabriel gabriel  17K jun  2 13:01 baitfile.fasta
-rw-rw-r-- 1 gabriel gabriel 529K jun  2 13:01 hashstat.bin
-rw-rw-r-- 1 gabriel gabriel  22M jun  2 13:02 Salternans-readpool-it2-se.fastq
-rw-rw-r-- 1 gabriel gabriel  23M jun  2 13:02 Salternans-readpool-it2.fastq
-rw-rw-r-- 1 gabriel gabriel  497 jun  2 13:02 manifest.se.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:04 Salternans-reference-se_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:04 backbone_it2_reference-pe.fna
-rw-rw-r-- 1 gabriel gabriel  492 jun  2 13:04 manifest.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:05 Salternans-reference_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans_reference-it2_noIUPAC.fasta
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$
```

Vamos precisar dessa montagem final. Use “cat” para imprimir essa sequência na tela.

A copie e salve em algum editor de texto no seu computador.

O nome desse arquivo pode ser “nome_da_espécie.mitobim.fa”

```
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$ ls -ltrh
total 46M
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:01 backbone_it2_initial_reference.fna
-rw-rw-r-- 1 gabriel gabriel  17K jun  2 13:01 baitfile.fasta
-rw-rw-r-- 1 gabriel gabriel 529K jun  2 13:01 hashstat.bin
-rw-rw-r-- 1 gabriel gabriel  22M jun  2 13:02 Salternans-readpool-it2-se.fastq
-rw-rw-r-- 1 gabriel gabriel  23M jun  2 13:02 Salternans-readpool-it2.fastq
-rw-rw-r-- 1 gabriel gabriel  497 jun  2 13:02 manifest.se.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:04 Salternans-reference-se_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:04 backbone_it2_reference-pe.fna
-rw-rw-r-- 1 gabriel gabriel  492 jun  2 13:04 manifest.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:05 Salternans_reference_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans_reference-it2_noIUPAC.fasta
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$ cat Salternans_reference-it2_noIUPAC.fasta
>SALTERNANS-NOVOPLASTY_bb_bb_bb_bb_bb_bb
cccacaatcggccatcctgcgattcatccgcctcggatgaagaatctccatcaccccatggcttcaagtattctgcgcaggtagaagtttgattaggtccttcatggcttccttctcgtagaacgtcatacg
gacgatgcgatacgggccgaggtactttttccggatcttcagccccggtccgaattgcgttcgcttaattgccacgatgtctcctatagtgtactggttagccggtttccgacgtcggttgaatgtcctctt
atgatctgctgcttggctgtttctctaaattctttttgtgagatgtagatcgttctgtgtcagatccagtagtagtcgttagactttgctcttttctacgattccgtctctccaaggacactttaacatcg
tctctttcaagcatcggtatttgtcttctagtaaacgtttctcgcagcgcactttctcgtataatctgtcttcacctaactcacgcattcttgttctttcattttgggtcactcctaggactatgcctagta
tattcctagcaatacattacactgccaacattgaaatagcttttaatagagtaagacatatttgtcgagatgtaaactatggatgactaatacgcacattacatgctaacggagcctcatttttttttatct
tcgaggaatatactacggatcatacaagctaatacatacatgaatagttggagtagttattttatttttagtaatagcaacggcctttataggatatgtcctcccttgaggacaaatatcattctgaggagc
ctttcagctgtcccctacttaggtactatgatagttcaatgagtttgaggaggatttgcagtagacaatgccaccttaactcgattctttacaatccactttctattaccatttattgtagctgcaatagtt
tacaccaaacagggtctaacaatcctctaggaataaatagtaatattgataaaattccttttcatccctacttctcatttaaggatctcataggattcatcattttattaataacattaacaacattatcac
aggagaccctgataacttcaccccagcaaatcctcttgtaaccccagtacacattcagcccgaatgatatttcctatttgcatatgcaatcctccgatcaattcctaataaactcggtggagtcattgccct
ttaataattctaccttttttacaatattaataaattccaatcaatgaaattttatcctatcaaccaattcctattttgatcattcttaatcattgtaatcctactcacatgaattggtatgcgaccagttgaa
gacaaattcttacagtcatatactttatatacttcatattaaacccactaatattcaaaatttgagactcaataatttactagttaatgaacttgtaaaagtatatgttttgaaaacataataaagaggtta
ctaaaattaattcactaaattaacaaagaataaaataaaaccctcaaccccaaatataaaaaaagataatttaaagccacaggtaaaaaactccttcaagctaaatatatcaacttatcataacgatatcga
atacaaaagtaaatgagataaacaccatccttaaataaaacattaaagaaaataaatcccccccctaaaaaaattacaacatataatatacttataaataaaattcttgagtactcagacaaaaaaattaaag
ctccacattaaatccagaaactaactcagactcaccttccgcaaaatcaaaaggagttcgattagtctcggctaatctagaaacaaatcaaataaatgctaaaggtaaagacaaaaaaattattcaaacata
tccagtaaacaaaaccctcaattaaaatcataaaagatattaaaattaaagccaaactaacttcataagaaatagtctgagctacagctcgcatacccccccaacatagaatagcttgaattagatgaccatccagcaatcattactgtat
```

Copy
Paste
Horizontal split
Vertical split
Close window
New workspace
Fullscreen
Search
Google
Bing
Rename title
Encoding
Custom commands
Remote management
Settings

Dentro dessa mesma pasta, haverá outra pasta chamada “nome_da_espécie-reference_assembly. Entre nela.

Obs: não entre na pasta de nome parecido: “nome_da_espécie-reference-**se**_assembly”

```
gabriel@Leveza: iteration2
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$ ls -ltrh
total 46M
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:01 backbone_it2_initial_reference.fna
-rw-rw-r-- 1 gabriel gabriel  17K jun  2 13:01 baitfile.fasta
-rw-rw-r-- 1 gabriel gabriel 529K jun  2 13:01 hashstat.bin
-rw-rw-r-- 1 gabriel gabriel  22M jun  2 13:02 Salternans-readpool-it2-se.fastq
-rw-rw-r-- 1 gabriel gabriel  23M jun  2 13:02 Salternans-readpool-it2.fastq
-rw-rw-r-- 1 gabriel gabriel  497 jun  2 13:02 manifest.se.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:04 Salternans-reference-se_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:04 backbone_it2_reference-pe.fna
-rw-rw-r-- 1 gabriel gabriel  492 jun  2 13:04 manifest.conf
drwxrwxr-x 6 gabriel gabriel 4,0K jun  2 13:05 Salternans-reference_assembly
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans_reference-it2_noIUPAC.fasta
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2$
```

Dentro da nova pasta, entre na pasta de resultados “_results”.

Daí, use “ls -ltrh” e identifique o arquivo com a extensão “.maf

```
gabriel@Leveza: Salternans-reference_assembly
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly$ ls -tlrh
total 16K
drwxrwxr-x 2 gabriel gabriel 4,0K jun  2 13:04 Salternans-reference_d_tmp
drwxrwxr-x 2 gabriel gabriel 4,0K jun  2 13:05 Salternans-reference_d_results
drwxrwxr-x 2 gabriel gabriel 4,0K jun  2 13:05 Salternans-reference_d_info
drwxrwxr-x 2 gabriel gabriel 4,0K jun  2 13:05 Salternans-reference_d_chkpt
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly$ cd Salternans-reference_d_results
```

```
gabriel@Leveza: Salternans-reference_d_results
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$ ls -tlrh
total 323M
-rw-rw-r-- 1 gabriel gabriel 199M jun  2 13:05 Salternans-reference_out.caf
-rw-rw-r-- 1 gabriel gabriel 124M jun  2 13:05 Salternans-reference_out.maf
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_reference.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_reference.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  184 jun  2 13:05 Salternans-reference_out.tcs
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out.wig
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$
```

E, dentro dessa pasta, precisaremos converter o arquivo “.maf” gerado em um outro tipo de arquivo, chamado “.ace”.

Para isso, precisaremos usar um módulo do MIRA, chamado mirabait. Em nossos servidores, ele está instalado na variável PATH (basicamente, um lugar da memória do computador que guarda atalhos da linha de comando), logo não precisamos digitar o caminho completo, só o nome do programa.

Digite: miraconvert -f maf -t ace -r C -r f arquivo.maf NOME_DA_ESPÉCIE.MITOBIM.NÚMERO_DA_ITERAÇÃO.NÚMERO_DE_READS_USADAS

Ex: miraconvert -f maf -t ace -r C -r f Salternans-reference_out.maf SALTERNANS.MITOBIM.IT2.10MILLION

```
gabriel@Leveza: Salternans-reference_d_results
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$ miraconvert -f maf -t ace -r C
 -r f Salternans-reference_out.maf SALTERNANS.MITOBIM.IT2.10MILLION
Loading from maf, saving to: ace
Loading MAF Salternans-reference_out.maf :
 [0%] ....|.... [10%] ....|.... [20%] ....|.... [30%] ....|.... [40%] ....|.... [50%] ....|.... [60%] ....|.... [70%] ....|.... [80%] ....|.... [90%] .
...|....Localtime: Sat Jun  2 18:09:50 2018

Generated 34887 unique DNA template ids for 67552 valid reads.
 [100%]
 done.

Data conversion process finished, no obvious errors encountered.
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$
```

E asim teremos o arquivo .ace que precisamos. Lembre-se que o nome desse arquivo deve seguir exatamente a estrutura:

NOME_DA_ESPÉCIE.MITOBIM.NÚMERO_DA_ITERAÇÃO.NÚMERO_DE_READS_USADAS.ace

No nosso exemplo: SALTERNANS.MITOBIM.IT2.10MILLION.ace

```
gabriel@Leveza: Salternans-reference_d_results
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$ ls -tlrh
total 625M
-rw-rw-r-- 1 gabriel gabriel 199M jun  2 13:05 Salternans-reference_out.caf
-rw-rw-r-- 1 gabriel gabriel 124M jun  2 13:05 Salternans-reference_out.maf
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_reference.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_reference.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  184 jun  2 13:05 Salternans-reference_out.tcs
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out.wig
-rw-rw-r-- 1 gabriel gabriel 303M jun  2 18:09 SALTERNANS.MITOBIM.IT2.10MILLION.ace
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$
```

Temos que compactar esse arquivo. Para isso, digite: “gzip *.ace”

E assim obteremos um arquivo bem menor, com extensão “.ace.gz”

```
gabriel@Leveza: Salternans-reference_d_results
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$ gzip *ace
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$ ls -tlrh
total 343M
-rw-rw-r-- 1 gabriel gabriel 199M jun  2 13:05 Salternans-reference_out.caf
-rw-rw-r-- 1 gabriel gabriel 124M jun  2 13:05 Salternans-reference_out.maf
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_AllStrains.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_AllStrains.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_Salternans.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_Salternans.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  57K jun  2 13:05 Salternans-reference_out_reference.padded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out_reference.padded.fasta
-rw-rw-r-- 1 gabriel gabriel  184 jun  2 13:05 Salternans-reference_out.tcs
-rw-rw-r-- 1 gabriel gabriel  53K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta.qual
-rw-rw-r-- 1 gabriel gabriel  18K jun  2 13:05 Salternans-reference_out_reference.unpadded.fasta
-rw-rw-r-- 1 gabriel gabriel  19K jun  2 13:05 Salternans-reference_out.wig
-rw-rw-r-- 1 gabriel gabriel  21M jun  2 18:09 SALTERNANS.MITOBIM.IT2.10MILLION.ace.gz
gabriel@Leveza:~/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results$
```

Agora, se você estiver utilizando um servidor, vamos ter que baixar esse arquivo no seu computador. Para isso, vamos usar o comando sftp (***S**ecure **F**ile **T**ransfer **P**rotocol*)

Abra um novo terminal e digite:

sftp nome_de_usuário@ip_do_servidor

E digite sua senha.

Use o comando “cd” para ir até a pasta na qual está o arquivo “.ace.gz”.

Para facilitar, use o “Tab” duplo para ir completando os caminhos e vendo os conteúdos das pastas.
O comando “ls -ltrh” também pode ajudar.

```
sftp> cd /home/gabriel/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results/
```

Por útlimo, use o camndo “get” para baixar o arquivo na sua máquina.

Digite: get nome_do_arquivo.ace.gz

```
sftp> get SALTERNANS.MITOBIM.IT2.10MILLION.ace.gz
Fetching /home/gabriel/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results/SALTERNANS.MITOBIM.IT2.10
MILLION.ace.gz to SALTERNANS.MITOBIM.IT2.10MILLION.ace.gz
/home/gabriel/100_MITO/NOVOPLASTY_Salternans/iteration2/Salternans-reference_assembly/Salternans-reference_d_results   0%   96KB  32.3KB/s   10:38 ETA
```

Vários programas podem abrir o arquivo ace, como por exemplo o Tablet (https://ics.hutton.ac.uk/tablet/download-tablet/).

Dependendo do resultado (se todo o mitogenoma possui uma alta cobertura, por exemplo), a mitocôndria já estará pronta e poderemos ir para a próxima etapa: a **anotação**.

Como a anotação é bem mais subjetiva que a montagem, ela não será retratada aqui.

Entretanto, o fasta da montagem pode ser submetido ao MITOS Web Server (http://mitos.bioinf.uni-leipzig.de/index.py) para uma anotação automática.

Essa anotação automática, após completada, te permitirá baixar um arquivo gff (generic feature format) que pode ser convertido usando meu script (mitos_to_artemis.py, encontrado em https://github.com/gavieira/python_bioinfo).

Essa conversão é necessária para se unir o arquivo fasta e o gff em um arquivo genbank através do software Artemis (http://sanger-pathogens.github.io/Artemis/Artemis/)