

***Loxodonta* Localizer (LL): on trimming sequences to be used as queries.**

TRIMMING SEQUENCES (quick guide)

Sequences must be trimmed to an exact 316 bp portion of the hypervariable region 1 of the control region of mitochondrial DNA. Here are the steps involved:

(1) You will need to use one of several programs that are useful for visualizing, checking, aligning and trimming DNA sequences. These include, but are not limited, to:

Geneious, <https://www.geneious.com>

Sequencher, <https://www.genecodes.com>

EMBL-EBI EMBOSS Matcher (free tool), <https://www.ebi.ac.uk/Tools/psa/>

(2) Once you have established that your sequences are of high quality, you can align your sequence(s) to a reference sequence from the LL, and then use the reference sequence as a guide to trim your generated sequence(s) to the correct region and size. Here is a reference sequence that can be copied and pasted into the software described above (or similar software), to use as a guide for trimming your sequences:

>LL001

```
TCAACCCGCTATGTATATCGTGCATTAATGCTTGTCCTCCCATACATAATGATATATATTACTAAC  
TATACTTAATCTTACATAGACCATACTATGTATAATCGTGCATCACATTATTTACCCCATGCTTA  
TAAGCAAGTACTGTTTAACTAATGTGTCAAGTCATATTCATGTAGATTCACAGGTCATGTTTTGG  
TCCATGGATATTGTTTACCCACGATAAACCATAGTCTTACATAGCACATTAAGCCCTTGATCGT  
GCATAGCACATCACTGAGAAATCTCTAGTCACCATGCATATCACCTCCAATGGTTG
```

(3) Using the reference sequence in (2) as a guide, trim your generated sequence.

(4) After trimming the sequences, it should be useful to first identify any identical mtDNA haplotype sequences across your trimmed sequences, and query each distinct haplotype only once. You can record information on all of the samples that share the same haplotype, and may choose to include this information in the first line of the FASTA format in the text line preceded by ">" as this information will be populated to the LL output listing.

(5) Use each trimmed sequence as a query on the home page of the LL. In FASTA format, any notes describing your query can be included in the first line, as text preceded by ">"

(6) Additional quality control step. If the LL output listing does not show an exact match to your query, the output listing can be used for quality control on any novel (previously unreported) sequence. In the output listing, use the "show alignment" option to identify the position of the character state that makes your novel sequence unique. Then verify in the raw data that this character state is real and does not reflect a technical error or ambiguity.

ADDITIONAL INFORMATION (details not generally needed to conduct trimming)

Ambiguous character states (such as N or R) are not allowed in the Loxodonta Localizer query sequence, which must only include the letters A, C, G, and T.

The query sequence corresponds to the same region of mtDNA reported by Johnson et al. 2007; BMC Evol Biol. 7:244.

Description of sequences immediately upstream and downstream of the sequences to be input as a query:

If you are using the primers CR-F1 and CR-R2, which are included in the *Loxodonta* Localizer “protocol for amplification and Sanger sequencing of DNA from dung or ivory,” then follow these guidelines: the query sequence will begin 94 bp downstream of the primer CR-F1 (or 93 bp if the sequence includes the deletion shown below) and end 112 bp upstream of the primer CR-R2. Below are shown only the 30 bp of sequences that immediately flank the appropriate 316 bp control sequence; these will be trimmed (along with the regions further upstream or downstream) in order to generate a sequence of the correct length. The numbers are positions in the elephant mitogenome reference sequence in GenBank with accession number AB443879 (Murata et al., 2009; *Gene* 441:176). Note that AB443879 has a deletion at position 15425 relative to many other elephant sequences, and this should be taken into account.

5' side

15396 15425
A T T C T T C T T A A A C T A T T C C C T G C A A G C A A A
G C G Δ
(Deletion)

3' side

15739 15770
T A C C T T A A C T A C C T A C C T C C G A G A A A C C A T
T

If using other primers or other methods to generate sequences, use the quick guide to trimming sequences at the top of this document.

Querying sequences that are identical across samples only once:

In confiscated batches of ivory, there may be samples from many different elephants that carry the same mtDNA sequences in the 316 bp region of hypervariable region 1 used by the *Loxodonta* Localizer. The user should, after trimming the sequences, compare all of the sequences to each other. For samples that carry the same 316 bp mtDNA control region sequence, these identical sequences should be queried against the LL only once, to avoid redundancy in the workload and in the LL output results.

When samples share identical sequences, text that identifies all of the identical samples can be included within a FASTA header line that can form part of the query against the *Loxodonta* Localizer database. The header line beginning with “>” can be followed by information on all samples that carry the query sequence, such as a description, sample number, haplotype ID, or other identifying information. All of the information included in FASTA format on the input page will be reproduced on the output page, which also displays a map and a listing of matching haplotypes, and which includes a timestamp that lists the date and time that the output was generated (in Greenwich Mean Time).

If a distinct novel sequence is detected in only one sample, the single nucleotide polymorphism(s) (SNPs) unique to that sample should be identified. This may be done using the “show alignment” option on the LL output page. As an extra quality control step, the sequence should be double-checked at the position(s) of the mismatching SNP(s), in order to ensure that any character state unique to the haplotype is not due to mis-calling of a base.