Sequence checksum

Introduction
A program to generate checksums for sequences. The checksums can be CRC32, CRC64-ISO, CRC64-ECMA-182, CRC32-CRC64, MD5, SHA1, SHA2-224-bit, SHA2-256-bit, SHA2-384-bit and SHA2-512-bit.

How to use this tool
Running a tool from the web form is a simple multiple steps process, starting at the top of the page and following the steps to the bottom.

Each tool has at least 2 steps, but most of them have more:

- The first steps are usually where the user sets the tool input (e.g. sequences, databases...)
- In the following steps, the user has the possibility to change the default tool parameters
- And finally, the last step is always the tool submission step, where the user can specify a title to be associated with the results and an email address for email notification. Using the submit button will effectively submit the information specified previously in the form to launch the tool on the server

Note that the parameters are validated prior to launching the tool on the server and in the event of a missing or wrong combination of parameters, the user will be notified directly in the form.

Step 1 - Input Sequences

Input Sequence

The input sequence can be entered directly into this form. Sequences can be in GCG, FASTA, EMBL (Nucleotide only), GenBank, PIR, NBRF, PHYLIP or UniProtKB/Swiss-Prot (Protein only) format. Partially formatted sequences are not accepted. [See example input formats]. Adding a return to the end of the sequence may help certain applications understand the input. Note that directly using data from word processors may yield unpredictable results as hidden/control characters may be present.

Sequence File Upload

The input sequence in any of the following formats (GCG, FASTA, EMBL (Nucleotide only), GenBank, PIR, NBRF, PHYLIP or UniProtKB/Swiss-Prot (Protein only)) can be uploaded and used as input. [See example input formats]. Word processor files may yield unpredictable results as hidden/control characters may be present in the files. It is best to save files with the Unix format option to avoid hidden Windows characters.

Sequence Type

Indicates if the sequences are protein or nucleotide (DNA/RNA).

<table>
<thead>
<tr>
<th>Type</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>protein</td>
</tr>
<tr>
<td>Nucleic Acid</td>
<td>dna</td>
</tr>
</tbody>
</table>

Default value is: Protein [protein]

Step 2 - Select Applications

Checksum Methods

A number of different checksum methods for sequences are launched.

<table>
<thead>
<tr>
<th>Method Name</th>
<th>Description</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC32</td>
<td>32-bit checksum using Cyclic Redundancy Check (CRC32).</td>
<td>crc</td>
</tr>
<tr>
<td>CRC64-ISO</td>
<td>64-bit checksum using the Cyclic Redundancy Check defined by ISO (CRC-64-ISO)</td>
<td>spcrc</td>
</tr>
<tr>
<td>CRC64-ECMA-182</td>
<td>64-bit checksum using the alternative Cyclic Redundancy Check defined in ECMA-182 (CRC-64-ECMA-182)</td>
<td>altcrc</td>
</tr>
<tr>
<td>CRC32-CRC64</td>
<td>Combination of CRC32 and CRC64 sequence checksums with the sequence length</td>
<td>cdigest</td>
</tr>
<tr>
<td>MD5</td>
<td>128-bit checksum using Message Digest 5 (MD5)</td>
<td>md</td>
</tr>
<tr>
<td>SHA1</td>
<td>160-bit checksum using Secure Hash Algorithm 1 (SHA-1)</td>
<td>shaa</td>
</tr>
</tbody>
</table>
SHA2-224  224-bit checksum using Secure Hash Algorithm 2 (SHA-2)  shab
SHA2-256  256-bit checksum using Secure Hash Algorithm 2 (SHA-2)  shac
SHA2-384  384-bit checksum using Secure Hash Algorithm 2 (SHA-2)  shad
SHA2-512  512-bit checksum using Secure Hash Algorithm 2 (SHA-2)  shae

Default value is: SHA2-512 [shae]

**Example output formats**

**Step 3 - Submission**

**Job title**

It’s possible to identify the tool result by giving it a name. This name will be associated to the results and might appear in some of the graphical representations of the results.

**Email Notification**

Running a tool is usually an interactive process, the results are delivered directly to the browser when they become available. Depending on the tool and its input parameters, this may take quite a long time. It’s possible to be notified by email when the job is finished by simply ticking the box “Be notified by email”. An email with a link to the results will be sent to the email address specified in the corresponding text box. Email notifications require valid email addresses.

**Email Address**

If email notification is requested, then a valid Internet email address in the form joe@example.org must be provided. This is not required when running the tool interactively (The results will be delivered to the browser window when they are ready).

**References**

_The EMBL-EBI bioinformatics web and programmatic tools framework._
(2015 April 06) Nucleic acids research 43 (W1) :W580-4
PMID: 25845596

_A new bioinformatics analysis tools framework at EMBL-EBI._
(2010 May 03) Nucleic acids research 38 (Web Server issue) :W695-9
PMID: 20439314

_Analysis Tool Web Services from the EMBL-EBI._
PMID: 23671338