**GeneTargeter: Automated in silico design for genome editing in the malaria parasite, Plasmodium falciparum**

Pablo Cárdenas 1, Lisl Y. Esherick1, Gaël Chambonnier1, Sumanta Dey1,2, Christopher V. Turlo1, Armiyaw Sebastian Nasamu1,3 and Jacquin C. Niles1,4\*

1Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave 02139 Cambridge, MA, USA

2Present address: Pfizer, Inc., Cambridge, MA, USA.

3Present address: Johns Hopkins Bloomberg School of Public Health, 615 Wolfe Street, Baltimore, MD 21205, USA.

4Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, Massachusetts 02142, USA

\*Correspondence**:** [jcniles@mit.edu](mailto:jcniles@mit.edu)

**Supplementary Material**

**Contents**

1. Results:
   1. Table S1: GeneTargeter DNA oligos used to assemble sample pSN054 constructs
   2. Table S2: Expected DNA digestion fragment sizes for sample pSN054 constructs
   3. Table S3: Expected PCR amplicon sizes for sample pSN054 constructs
2. Methods:
   1. Figure S1: GeneTargeter software architecture
   2. Table S4: GeneTargeter additional optional parameters

**Table S1**: GeneTargeter DNA oligos used to assemble sample pSN054 constructs

|  |  |
| --- | --- |
| **Primer Name** | **Primer Sequence** |
| PF3D7\_1401800 (CK) LHR fw | ttcaaacttcattgactgtgccggccggccaaaactccttgtgtatttaaaatgatggat |
| PF3D7\_1401800 (CK) LHR rv | cgcttgcggctgtgctgtcgtctagatatttggatagataagtcgttacaaaaagtattc |
| PF3D7\_1401800 (CK) RHR fw | tgtacggtacaaacccggaattcgagctcggtcaaaaaaaacagaaaaaatttacattct |
| PF3D7\_1401800 (CK) RHR rv | gggtattagacctagggataacagggtaataaaagaaaagaaaaagaaggaaaaaaatta |
| PF3D7\_1124600 (EK) LHR fw | ttcaaacttcattgactgtgccggccggcctcacccgtggttctttgtcattgtgatttg |
| PF3D7\_1124600 (EK) LHR rv | agggttagggataggcttacccgcgatcgcgtttttttccaatttgctcctaaattttac |
| PF3D7\_1124600 (EK) RHR fw | gtacggtacaaacccggaattcgagctcggtgctaaaaaaattatgaacagttcattatt |
| PF3D7\_1124600 (EK) RHR rv | gggtattagacctagggataacagggtaatgaagggaagggaagggaaagcaaagcaaag |
| PF3D7\_0910000 (PfSET4) LHR fw | ttcaaacttcattgactgtgccggccggccatattttaaatgtcaaaaatgcagaggacg |
| PF3D7\_0910000 (PfSET4) LHR rv | accaacgggaccaagtccgtgtgtatcgcctggggacattgaatttgaataaaatagaaa |
| PF3D7\_0910000 (PfSET4) RHR fw | gtacggtacaaacccggaattcgagctcggttatatcccttctcaaaagatgagatgttt |
| PF3D7\_0910000 (PfSET4) RHR rv | gggtattagacctagggataacagggtaatagctgaaatagcagcagtattatcatgttt |
| PF3D7\_1355300 (PfSET6) LHR fw | ttcaaacttcattgactgtgccggccggccgattagtagacaaagaaaaagaaaccttta |
| PF3D7\_1355300 (PfSET6) LHR rv | tccaagtccttgtatatcaagctgtccggaccatacgttttaataatatttttttttgct |
| PF3D7\_1355300 (PfSET6) RHR fw | gtacggtacaaacccggaattcgagctcggaaaaagggaaataaaataaaaaaagaaaca |
| PF3D7\_1355300 (PfSET6) RHR rv | gggtattagacctagggataacagggtaattttccttaaaaacaataatgtggaacacat |
| PF3D7\_1115200 (PfSET7) LHR fw | tttcaaacttcattgactgtgccggccggccacatctcaatgggaaaaagcaatatacta |
| PF3D7\_1115200 (PfSET7) LHR rv | cgtaaggttgtcaatgcggtagtacttgttcaatgaatctcttaaataatcatgagagta |
| PF3D7\_1115200 (PfSET7) RHR fw | tgtacggtacaaacccggaattcgagctcggggtggaaaatgaaattaaagaatttcaaa |
| PF3D7\_1115200 (PfSET7) RHR rv | gggtattagacctagggataacagggtaataataaatttttgaatatcattattccggtt |
| PF3D7\_1304600 (PfLSMT) LHR fw | ttcaaacttcattgactgtgccggccggccttgtaacagttgtaaatatgtaggatgttc |
| PF3D7\_1304600 (PfLSMT) LHR rv | ccaaagccgttcacgtgtacgttaatccacttttcaatatagttgtatacatgatttgtc |
| PF3D7\_1304600 (PfLSMT) RHR fw | gtacggtacaaacccggaattcgagctcgggttattataaagtttaaaggaattacacaa |
| PF3D7\_1304600 (PfLSMT) RHR rv | gggtattagacctagggataacagggtaattataaaaaaatacaaacatgtgcttcctcc |
| PF3D7\_0922200 (SAMS) LHR fw | ttcaaacttcattgactgtgccggccggccatttcaactcaacatgctgaagatataaaa |
| PF3D7\_0922200 (SAMS) LHR rv | ctaggtcgtagtccgtgtagcccgtgcttacggtaccataagaattaacattaagtgata |
| PF3D7\_0922200 (SAMS) RHR fw | gtacggtacaaacccggaattcgagctcggaaaaaaaaacatcatattataattaagcgt |
| PF3D7\_0922200 (SAMS) RHR rv | gggtattagacctagggataacagggtaataaaaaaatagaaacaaggtcaaatataaac |

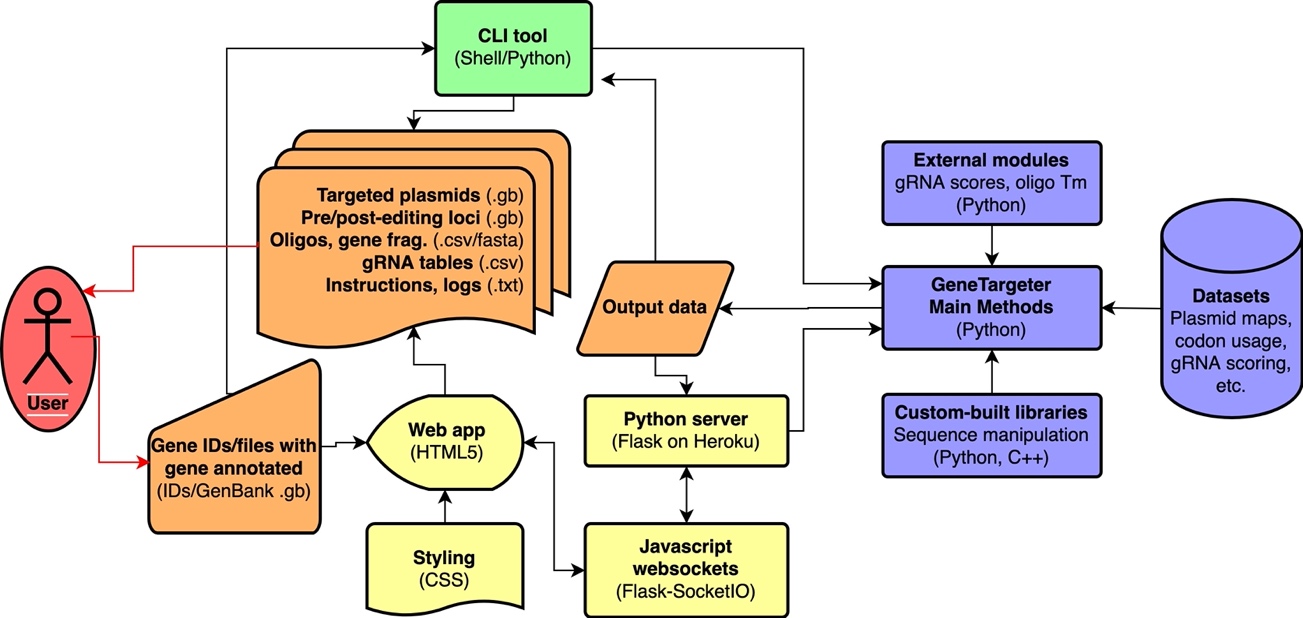
**Table S2**: Expected DNA digestion fragment sizes for sample pSN054 constructs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Target** | **Restriction digest expexted sizes (bp)** | | | |
| AscI + I-SceI | FseI + AsiSI | ApaI + XmaI | ApaI + XbaI |
| PF3D7\_1401800 (CK) | 17822, 42, **964**, 2094 | 10358, **649**, 9915 | 11125, 73, **861**, 8863 | 10207, 558, 48, 385, **1742**, 3996, 3986 |
| PF3D7\_1124600 (EK) | 17591, 42, **654**, 2094 | 10358, **418**, 9605 | 10894, 73, **861**, 8553 | 10207, 760, **1742**, 3996, 3676 |
| PF3D7\_0910000 (PfSET4) | 17851, 42, **654**, 2094 | 10358, **678**, 9605 | 11154, 73, **861**, 8553 | 10207, 1020, **1742**, 3996, 3676 |
| PF3D7\_1355300 (PfSET6) | 17635, 42, **813**, 2094 | 10358, **462**, 9764 | 10938, 73, **861**, 8712 | 10207, 804, **1742**, 3996, 3835 |
| PF3D7\_1115200 (PfSET7) | 17650, 42, **728**, 2094 | 10358, **477**, 9679 | 10953, 73, **861**, 8627 | 10207, 819, **1742**, 3996, 3750 |
| PF3D7\_1304600 (PfLSMT) | 17957, 42, **654**, 2094 | 10358, **784**, 9605 | 11260, 73, **861**, 8553 | 10207, 445, 681, **1742**, 3996, 3676 |
| PF3D7\_0922200 (SAMS) | 17786, 42, **755**, 2094 | 10358, **613**, 9706 | 10822, 267, 73, **861**, 8654 | 10207, 955, **1742**, 3996, 3777 |

**Table S3**: Expected PCR amplicon sizes for sample pSN054 constructs

|  |  |  |
| --- | --- | --- |
| **Target** | **Homology Region Amplicon size (bp)** | |
| **LHR** | **RHR** |
| PF3D7\_1401800 (CK) | 460 | 772 |
| PF3D7\_1124600 (EK) | 471 | 461 |
| PF3D7\_0910000 (PfSET4) | 460 | 461 |
| PF3D7\_1355300 (PfSET6) | 460 | 620 |
| PF3D7\_1115200 (PfSET7) | 474 | 536 |
| PF3D7\_1304600 (PfLSMT) | 461 | 462 |
| PF3D7\_0922200 (SAMS) | 460 | 562 |

**Figure S1**: GeneTargeter software architecture



**Figure S1** GeneTargeter combines multiple modules to allow flexibility in usage. The user provides GenBank gene files which may be processed through the CLI tool or the Web application, both of which rely on a series of Python modules and databases to produce the output files for the gene(s) provided.

**Table S4**: GeneTargeter additional optional parameters

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Default value** | **Comments** |
| Use manual sgRNA, LHR, and/or RHR annotations | False | Instructs GeneTargeter whether to use manually-selected annotations within the input file for the design process (True/False). |
| Treat as non-coding | False | Instructs GeneTargeter to search for edits exclusively outside of the gene sequence; if Auto, GeneTargeter evaluates whether gene is non-coding based on absence of stop codon (True/False/Auto). |
| Plasmid type | pSN054\_V5 | Defines which plasmid type to use (’pSN054V5’, ’pSN054’, ’pSN150’, ’pSN150-KO’ for 3’, 5’, and for gene knockouts, respectively; also ’pSN150-Ter’, ’pSN150-KO-Ter’ with modified T7 terminator with no Nhe1 sequence). |
| Use HA tags | True | Adds or removes a hemagglutinin (HA) tag fusion to the target (True/False). |
| LHR and RHR length (minimum, preferred, maxi-mum) | 400 bp, 500bp, 750 bp | GeneTargeter will optimize over the range given by the minimum and maximum values if the preferred one possesses unsatisfactory melting temperature. |
| LHR and RHR start and end optimization range | –20 bp to 20bp | After choosing valid start and end positions, the software searches the given range around each position for the value with the highest melting temperature. |
| LHR and RHR end analysis length | 40 bp | Length of sequences at the ends (start and finish) of the LHR and RHR taken into account for melting temperature analysis. |
| LHR and RHR end minimum melting temperature | 55 °C | Minimum melting temperature of sequences at the start and end of the LHR and RHR. |
| LHR maximum distance from sgRNA | 500 bp | Maximum distance the LHR can end from the most upstream sgRNA. GeneTargeter will use this to restrain its LHR end point search range. |
| RHR maximum distance from gene | 500 bp | Maximum distance the RHR can start from the end of the gene. GeneTargeter will use this to restrain its RHR start point search range. |
| Minimum sgRNA GC content | 25% | Minimum GC content a potential sgRNA must have to be considered minimally acceptable. |
| CRISPR cutting enzyme | Cas9 | Type of CRISPR enzyme being used, determines PAM location as 3’ or 5’, for Cas9 and Cas12a, respectively (only accepts “cas9” or “cpf1”). |
| CRISPR enzyme PAM sequence | NGG | PAM sequence used by CRISPR enzyme, determines search motif used by GeneTargeter to locate potential sgRNAs. |
| On-target scoring method | Azimuth | Algorithm used to determine on-target scores for potential sgRNAs (only accepts “azimuth”, “ruleset2”, and “cindel”). |
| Minimum on-target score | 35% | Minimum on-target score for potential sgRNAs to be considered minimally acceptable. |
| Off-target scoring method | CFD | Algorithm used to determine off-target scores for potential sgRNAs (only accepts “cfd”, “hsu”, and “hsu unweighted”). |
| Maximum pairwise off-targetscore | 50% | Maximum off-target score a single hit can have for potential sgRNAs to be considered minimally acceptable. |
| Minimum total off-target score | 20% | Minimum aggregated off-target score for potential sgRNAs to be considered minimally acceptable. |
| Gibson homology range length (minimum, preferred, maximum) | 30 bp, 40 bp, 50 bp | GeneTargeter will optimize over the range given by the minimum and maximum values if the preferred one possesses unsatisfactory melting temperature. Each primer will be designed with approximately twice the homologous region size. |
| Gibson primer minimum melting temperature | 65°C | Minimum melting temperature of Gib-son primers. |
| Gibson primer maximum melting temperature difference | 5°C | Maximum melting temperature differ-ence of a pair of Gibson primers. |
| Codon optimization organism | T. gondii | Codon frequency table used for optimization (can be ’P. falciparum 3D7’,’P. vivax’, ’T. gondii’, ’E. coli K12’, ’S. cerevisiae’, ’H. sapiens’, ’R. norvegicus’, ’scramble’). |
| Use codon sampling instead of CAI maximization | False | GeneTargeter uses a deterministic CAI maximization algorithm [1] for codon optimization by default. To use a probabilistic codon sampling approach, set this argument to True (True/False). |
| Minimum gene fragment size | 250 bp | Minimum synthesis size of gene fragments. |
| Extended gene fragments | True | GeneTargeter extends a recoded region to the minimum gene fragment size if this requirement is set to True, otherwise, anneal-extension oligonucleotides or overhangs are used. |
| Consolidate output files | False | Aggregates all gene outputs into a single file for each output type, containing all gene outputs of that given file type; oligos are sorted by purpose and orientation before gene (True, False). |
| Prefix | \*None\* | If not \*None\*, adds this prefix to all file, oligo, DNA fragment, and Gen-Bank names. |
| Prefix number | −1 | If prefix is not \*None\* and a number >−1 is given here, numbers each gene processed consecutively starting at the given value; adds number to prefix on all file, oligo, DNA fragment, and GenBank names (integer) |