2015

Enabling broad use of genome modification technologies to solve real world problems: a specific application in peanut

Vincent Antonio Brazelton Jr.
Iowa State University

Enabling broad use of genome modification technologies to solve real world problems:

A specific application in peanut

by

Vincent A. Brazelton Jr.

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Interdepartmental Genetics and Genomics

Program of Study Committee:
Carolyn Lawrence-Dill, Major Professor
Steven Canon, Co-Major Professor
Jeffrey Wolt, Co-Major Professor

Iowa State University

Ames, Iowa

2015

DEDICATION

This manuscript is respectfully dedicated to my wife and children. You may never know how much of an inspiration you have been on my life. I have lived everyday with you in mind and it has made all the difference. I love you all.

-Vincent A. Brazelton Jr. M.S.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First and foremost, I would like to thank Jehovah Jireh for his unwavering favor and grace. He is the strength of my life and without him I am nothing.

I would also like to thank my committee chair, Carolyn Lawrence, and my committee members, Steven Cannon, and Jeffrey Wolt for their guidance and support throughout the course of this research.

To my family, close friends, and kinfolk, . . . you mean the world to me! Thank you for your constant prayers, your well wishes, your thoughts, your money, your encouraging words, your advice and most importantly your love. You are my driving force on my darkest days and without you none of this would be possible. I love you all and I look forward to what God has for us in the future.

In addition, I would also like to thank the Black Graduate Student Association (aka my extended family) for making my time at Iowa State University a wonderful experience. You have supported me through the harsh winters, the late night study sessions and the countless hours of trying to figure out why I decided to attend graduate school. It is extremely hard to believe none of you even existed three years ago! Yet today you are some of the closest friends and family I have.

I would also like to offer my appreciation to the Ice Cold Brothers of the Zeta Kappa Lambda Chapter of Alpha Phi Alpha Fraternity Incorporated. I am constantly in awe just to be present in a room with you brothers. Your dedication to the better making of men has made all the difference in my time here at Iowa State University, and I am honored to be counted amongst your ranks.

Finally, I would like to thank Ellen Tisdale M.S. I am positive that without you here at Iowa State to hold my hand, I would not have been able to survive. You are by far the most beautiful soul I have ever met and I am blessed that God allowed our paths to cross. You pour unselfishly of your time, love, and energy into all of my dreams. You are my biggest source of inspiration, my motivational speaker, my most outspoken critic, my confidant, my gladiator, the real MVP, (insert other words or phrases that mean I like you and stuff), but most importantly you are tha homie! I love you and I am excited to spend the remainder of my life with you.

CHAPTER I

INTRODUCTION

Biotechnology refers to a collection of techniques that uses living organisms or their products to make or modify a product, to improve plants or animals, or to develop microorganisms for specific uses (Abelson, 1992; Glick and Pasternak, 2003; Phillips, 2002). ~~A~~s new techniques emerge, the use of biotechnology to address emerging needs expands across various disciplines; including forensic science, pharmacology, medical technology, and agriculture.

Commercial adoption of plant biotechnology techniques has resulted in numerous agronomic traits including: increased nutrient-use efficiency, increased pest and pathogen resistance, and enhanced herbicide tolerance (James, 2014). In 1999, glyphosate-resistant varieties of soybean, cotton, and maize were released. Since that time, multiple techniques have been applied to addressing various agronomic issues. Today, global production of biotech crops totals more than 448 million acres of commercial farmland (James, 2014).

The potential for biotech to positively impact human and animal health, agricultural sustainability, and land use efficiency are well documented (Ciftci, 2000; Datta, 2013; Gersbach et al., 2007; Radakovits et al., 2010). However, there are several obstacles to the development and deployment of crops with biotech-derived, value-added traits: (1) most crop genomes are large and repetitive, which makes it computationally difficult to select targets for gene editing; (2) insufficient facilities (including greenhouse space) limit the scale of experiments that can be pursued; and (3) prohibitive costs associated with regulatory and licensing limit the number of biotech traits brought to market. In addition, the public perception of biotech as it pertains to food can be described as mixed at best. Many people are concerned that biotech crops – and

food products derived from them - are unsafe, a perspective that is not supported by scientific evidence (Blancke et al., 2015; Boudry et al., 2014; Giddings, 2015).

My research focuses on the first major obstacle: creating efficient genome analysis tools to enable development of improved lines and cultivars. In Chapter 2, I describe the current set of design tools for CRISPR-based gene editing, with emphasis on the CRISPR Genome Analysis Tool (CGAT) I co-developed with programmer Scott Zarecor. In Chapter 3, I describe my efforts to design functional knockouts of the major peanut allergen Ara h 1 using the CRISPR/Cas system for gene editing. This exercise is one example of how biotechnology can be used to develop traits to positively impact human health and food safety. In Chapter 4, I describe these projects' potential for positive impact and discuss whether and how they relate to public perception of biotech in agriculture.

CHAPTER II

A GUIDE TO CRISPR GENOME EDITING TOOLS

An invited article for: *GM Crops & Food: Biotechnology in Agriculture and the Food Chain* special issue on '*Gene Editing for Crop Improvement*'

Authors: Vincent A. Brazelton, Jr.[1,2†], Scott Zarecor[3†], David A. Wright[3], Yuan Wang[4,5], Jie Liu[4,5], Keting Chen[4,5], Bing Yang[3], and Carolyn J. Lawrence-Dill[1,2,3,4*]

[1]Interdepartmental Genetics and Genomics Program, Iowa State University, Ames, IA 50011

[2]Department of Agronomy, Iowa State University, Ames, IA 50011

[3]Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, 50011

[4]Interdepartmental Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011

[5]Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50011

†Contributed equally to the development of this project and manuscript

*Corresponding author

ABSTRACT

Targeted gene editing is now possible in nearly any organism and is widely acknowledged as a biotech game-changer. Among available gene editing techniques, the CRISPR-Cas9 system is the current favorite because it has been shown to work in many species, does not necessarily result in the addition of foreign DNA at the target site, and follows a set of simple design rules for target selection. This has resulted in the availability of an array of CRISPR design tools that vary in design specifications and parameter choices, available genomes, graphical visualization, and downstream analysis functionality. To help researchers choose a tool that best suits their specific research needs, we review the functionality of various CRISPR design tools including our own, the CRISPR Genome Analysis Tool (CGAT; http://cropbioengineering.iastate.edu/cgat).

INTRODUCTION

Early in the 20th century Muller showed that X-rays cause genetic mutations in Drosophila (Muller, 1927). Likewise, Stadler showed the mutational effects of X-rays on barley and maize (Stadler, 1928, 1944) which paved the way for researchers to broadly use mutagens such as X-rays and chemical agents to induce random genetic changes.

However, those methods yielded many mutations that had to be sorted out over generations to isolate the one responsible for causing changes to specific phenotypes/traits of interest. More recently, basic research to understand the processes underlying natural chromosomal recombination, microbial immune and virulence responses, and DNA binding domains led to discoveries that have made possible the development of *targeted* genome editing techniques that pair sequence-specific DNA binding proteins with

enzymes that cleave DNA (reviewed in (Wright et al., 2014). Development of these methods led to the realization that a RNA directed bacterial immune system could also be developed into an effective genome editing tool. Now three major systems for genome editing exist: Zinc Finger Nucleases (ZFNs), TAL Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs)-CRISPR associated proteins 9 (CRISPR-Cas9; reviewed in (Peng et al., 2014).

Zinc finger proteins are classified into distinct families based on specific structural motifs. Shared among all are DNA binding domains along with one or more zinc ion(s) that serve to stabilize the fold (Klug, 2010). Early NMR spectroscopy experiments revealed that the Cys2His2 zinc finger binding domain in the *Xenopus* transcription factor IIA is comprised of a 30 amino acid repeat sequence with conserved ββα secondary structure (Ruiz i Altaba et al., 1987). This architecture allows amino acids on the surface of the α-helix to interact with specific major groove nucleotides, thus conferring specificity for particular double- stranded DNA sequences (Beerli and Barbas, 2002); (Gaj et al., 2013); (Lee et al., 1989). It was later found that by changing amino acids in the α-helix, DNA binding specificity and affinity could be altered. Engineered zinc fingers were combined with the DNA cleavage domain of FokI, a type IIs restriction endonuclease, to form zinc finger nucleases (ZFNs), which were show to make specific targeted double-strand breaks in DNA. Induction of DNA damage triggers the cellular repair pathway via error-prone non-homologous end joining or template mediated homology directed repair thus giving limited control over the repair process in a targeted manner (Lieber, 2010). Non-homologous end joining can create loss- of-function mutations due to insertions, deletions, or rearrangements

whereas homology directed repair can create a precise mutation in the presence of a specific DNA template (Bogdanove, 2014; Lieber, 2010). ZFNs are known to cleave at off-target sites. This hampers their use and has been shown to cause cellular toxicity (Gaj et al., 2013); (Jiang et al., 2013)). ZFNs are also difficult (and costly) to design and construct with variable rates of success (reviewed in (Gaj et al., 2013; Jiang et al., 2013)).

Transcription activator-like effector (TALE; also called TAL effector) proteins are major components of the type III secretion system conferring pathogenicity in the Gram negative bacteria *Xanthomonas* (Bogdanove, 2014; Chen et al., 2010; Mak et al., 2013). Of the more than 30 families of bacterial effector proteins, TALEs are unique in their ability to distinguish specific DNA sequences via a central repetitive 34 amino acid DNA binding motif (Moscou and Bogdanove, 2009; Streubel et al., 2012). The repeat variable di-amino acids (RVDs) at positions 12 and 13 determine overall specificity and affinity for specific nucleotides in a target sequence. When coupled with the nuclease domain of Fok-I, TALE nucleases (TALENs) emerged as a novel genome-editing tool (Christian et al., 2010); (Li et al., 2011); (Boch et al., 2009).

Compared to ZFNs, TALEN-assisted genome editing has significantly reduced toxicity due to off-target effects; however, construct design complexity due to specific requirements in base composition coupled with a lack of support for the TALEN lentiviral delivery systems (reviewed in (Gaj et al., 2013); (Holkers et al., 2013) have held back broad adoption and use of TALENs (Sander and Joung, 2014).

The difficulties of both ZFN and TALEN techniques lie in designing and validating proteins that recognize specific DNA sequences. In contrast, the CRISPR

system is RNA- mediated. CRISPR was originally identified as a defense mechanism that provides bacterial adaptive immunity to a wide range of potential pathogens (). There are three major classes (types I, II, III) and ten subclasses of CRISPRs based on which specific CRISPR-associated (Cas) proteins and non-coding RNA species are involved (Carte et al., 2014; Makarova et al., 2011). The type II CRISPR-Cas9 system has been co-opted for gene editing.

The native CRISPR-Cas9 system (Figure 1; next page) is comprised of three distinct architectural components: a small non-coding transactivating CRISPR RNA (tracrRNA), an operon that encodes the Cas proteins, and a repeat array encompassing crRNA units comprised of a 5' 20-nucleotide targeting sequence and a 19-22 nucleotide repeat sequence (referred to as spacers; (Deltcheva et al., 2011). Multiple studies suggest that Cas9 endonuclease activity requires a highly conserved 3' three nucleotide protospacer adjacent motif (PAM) directly preceding the target sequence (Jiang et al., 2013; Zhang et al., 2014). PAM sequence composition is highly diverse depending on the CRISPR type/subtype with NGG representing the most effective trinucleotide for the CRISPR-Cas9 system of *Streptococcus pyogenes* (Zhang et al., 2014).

The native CRISPR-Cas9 genome editing mechanism is broken into a 3 processes: acquisition, expression, and interference (Carte et al., 2014; Makarova et al., 2011). Upon host infection, exogenous genetic elements are incorporated into the CRISPR locus (acquisition phase). These repeat sequences are then transcribed into noncoding precursor CRISPR RNAs (pre-crRNAs; expression phase). The Cas9 nuclease uses these guide RNA sequences to cleave invading plasmids or phage molecules including any double stranded DNA matching the CRISPR RNAs (interference). Double strand DNA breaks are repaired

via non-homologous end joining or homology directed repair *in vivo*, frequently leading to errors or elimination of invading DNA.

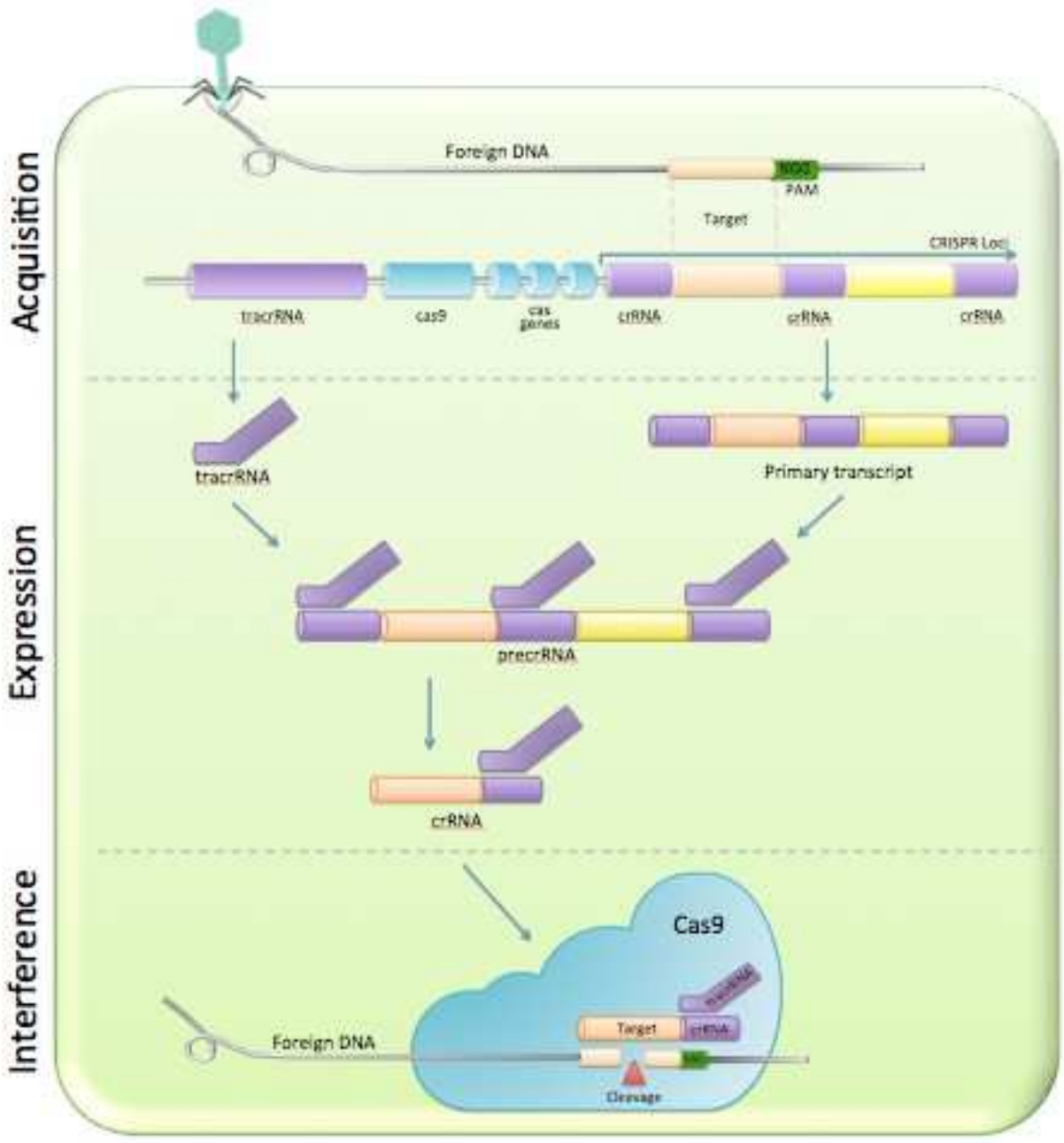


FIGURE 1: The CRISPR-Cas adaptive immune system. Three processes underlie the system, acquisition, expression, and interference. Foreign DNA is shown entering the cell. During acquisition, target DNA (beige; next to the PAM sequence shown in green) is incorporated into the CRISPR locus. Expression involves transcribed target DNA into noncoding pre-crRNAs to which tracrRNAs attach. During interference the Cas9 endonuclease uses these sequences to target foreign DNA for cleavage.

To simplify the system for targeted mutation, researchers combined the endogenous tracrRNA and crRNA to produce effective single guide RNA (sgRNA) constructs with unique restriction sites for targeting oligo insertion. The broad applicability of CRISPR to gene editing in diverse species coupled with simple design rules has resulted in the development of myriad bioinformatics tools that aim to identify potential sgRNA target sites in genomes of interest. Although multiple CRISPR sequence design tools already exist, they are not all the same. Some are user friendly, others are more difficult to use. Some are available via web servers, others are not available online. Many perform only a few steps in a full computational analysis and design pipeline, and deliver results that are voluminous with no mechanism to sort. In addition, few computational tools are solely dedicated to plant-based genomic analysis and fewer tools have been subjected to peer-review. To help researchers choose a tool that best suits their specific research needs, we compared the functionality of various CRISPR design software including our own, CGAT the CRISPR Genome Analysis Tool.

CRISPR COMPUTATIONAL RESOURCES COMPARISON

Of the available CRISPR tools we evaluated (see Table 1), there are two major classes: those that enable researchers to query experimentally validated sgRNAs for which genetic stocks are available, and those that predict potential CRISPR targets in a given sequence. At the time of this writing, the only tool we find that is in the former category is CrisprGE, though we anticipate that other tools will develop such resources in the very near future. For the remainder of this discussion, we focus on tools that can be used to predict potential CRISPR targets given an input sequence.

TABLE 1. CRISPR tool and resources examined.

Tool Name	Species	Publication	Web Address
Cas9 - Design	vertebrates, invertebrates, plants	Ma et al., 2013	http://cas9.cbi.pku.edu.cn/
CCTop	vertebrates, invertebrates, plants,	Stemmer et al., 2015	http://crispr.cos.uni-heidelberg.de/
CGAT	plants	This paper	http://cbc.gdcb.iastate.edu/cgat/
Chop-Chop	vertebrates, invertebrates, plants	Montague et al., 2014	https://chopchop.rc.fas.harvard.edu/
COSMID	vertebrates, invertebrates	Cradick et al., 2014	https://crispr.bme.gatech.edu/
CRISPR design	vertebrates, invertebrates, arabidopsis	N/A	http://crispr.mit.edu/
CRISPRdirect	vertebrates, invertebrates, fungi	Naito et al., 2014	http://crispr.dbcls.jp/
Crispr Finder	vertebrates invertebrates fungi	Grissa et al., 2007	http://crispr.u-psud.fr/Server/
CrisprGE*	various: plants, animals, fungi, prokaryotes, protists	Kaur et al., 2015	http://crdd.osdd.net/servers/crisprge/
CRISPR Multitargeter	vertebrates, invertebrates, plants	Prykhozhij et al., 2015	http://www.multicrispr.net/
Crispr-P	plants	Lei et al., 2014	http://cbi.hzau.edu.cn/crispr/
CRISPRseek	vertebrates, invertebrates, fungi, plants, protists	Zhu et al., 2014	http://www.bioconductor.org/packages/release/bioc/html/CRISPRseek.html
CROP-IT	vertebrates: mouse and human	Singh et al., 2015	http://cheetah.bioch.virginia.edu/AdliLab/CROP-IT/homepage.html
E-crisp	vertebrates, invertebrates, plants, fungi, protists	Heigwer et al., 2014	http://www.e-crisp.org/E-CRISP/
flyCRISPR	invertebrates	Gratz et al., 2014	http://flycrispr.molbio.wisc.edu/
GT-SCAN	vertebrates, invertebrates, plants, fungi	O'Brien and Bailey, 2014	http://flycrispr.molbio.wisc.edu/
sgRNAcas9	vertebrates, invertebrates	Xie et al., 2014	http://www.biootools.com/col.jsp?id=140
SSFinder	N/A	Upadhyay and Sharma, 2014	https://code.google.com/p/ssfinder/

*queries sgRNA sequences against experimentally validated sgRNAs for which genetic stocks are available

Multiple computational tools are available to aid in the prediction and design of CRISPR constructs to target specific genomic loci. Tools are classified based on whether

they are available online via web server, ability to search by gene name, options to use alternate PAM sequences, options to predict off-targets (by genomic sequence similarity), whether identified target lists can be sorted and/or ranked, and whether all of these functions are aggregated within a single, all-in-one pipeline. Here we specifically highlight the functionality of 17 non-commercial CRISPR design tools and report on their comparative functionality (Table 2).

TABLE 2. Comparison of CRISPR tool functionalities.

Tool Name	Web Server	Search by Gene Name	Alternate PAM Sequence	Predicts Off-targets	Ranks Output	All in One Tool
Cas9-Design	✓	✗	✗	✓	✗	✓
CCTop	✓	✗	✓	✓	✗	✓
CGAT	✓	✓	✓	✓	✓	✓
Chop-Chop	✓	✓	✓	✓	✓	✓
COSMID	✓	✗	✓	✓	✓	✓
CRISPR design	✓	✗	✓	✓	✓	✓
CRISPRdirect	✓	✓	✓	✓	✓	✓
Crispr Finder	✓	✗	✗	✓	✗	✗
CRISPR Multitargeter	✓	✗	✓	✓	✗	✗
Crispr-P	✓	✓	✓	✓	✓	✓
CRISPRseek	✗	✗	✓	✓	✗	✓
CROP-IT	✓	✓	✓	✓	✓	✗
E-crisp	✓	✗	✓	✓	✗	✗
flyCRISPR	✓	✗	✓	✗	✗	✓
GT-SCAN	✓	✗	✓	✓	✗	✓
sgRNAcas9	✗	✗	✓	✓	✗	✗
SSFinder	✗	✗	✗	✗	✗	✗

CRISPRseek, sgRNAcas9 and SSFinder are only available as stand-alone systems and require installation and configuration. CRISPR target sequences are identified and evaluated based on user input. These tools are best suited for users with some technical experience. Traditional experimental labs looking to quickly parse an input for possible CRISPR

targets/off-targets in their species of interest might be better served to try the tools accessible online via web server.

Beyond databases of validated CRISPR constructs and tools that must be downloaded and installed, myriad online tools exist that allow users to quickly parse an input to predict putative CRISPR targets. Tools in this category tend to allow the greatest amount of user flexibility in terms of sgRNA design criteria. As the CRISPR system continues to improve, specifications such as the ability to search non-canonical PAM sequences, an option to designate promoter-specific bases preceding the seed sequence, and improved prioritization for potential targets will provide the greatest expansion in utility across a multitude of genomes and cell types.

A major concern with targeted nuclease technology is the potential for off-target cleavage and associated toxicity. With this in mind, many tools check the rest of a genome for additional matches to predicted target sequences. Even more sophisticated tools produce a ranked output of CRISPR targets by interpreting off-target scores as a function of the overall sgRNA score.

Only CGAT, Crispr-P, Chop-Chop and CRISPRdirect offer access online, enable search by gene name, allow the use of non-canonical PAM sequences, predict off-targets, enable ranking of identified targets, and contains all of these functionalities within a single pipeline. Here we describe the functionality of CGAT and demonstrate its functionality as a specific example that shows how such tools work.

MATERIALS AND METHODS

CGAT is built upon a variety of technologies. PostgreSQL 9.3 (http://www.postgresql.org/) is the relational database system (RDBMS). For data retrieval, CGAT makes use of PostgreSQL's procedural language extensibility with portions of the database query logic written in PL/Python (http://www.postgresql.org/docs/9.3/static/plpython.html). The current version of the parser that processes genomic fasta-formatted files into relational database tables is written in the Go programming language (version 1.4.2) (https://golang.org/).

The website itself is written in Python 2.7.x using the 1.8.2 version of the Django framework (https://www.djangoproject.com/). Finally, the client-side functionality of the tool is written in Javascript using the 1.3.9 version of the AngularJS framework (https://angularjs.org/). Code is available online at https://github.com/ISU-Crop-Bioengineering-Consortium/crispr. While the above technology stack is relatively stable, version numbers of discrete pieces of the stack are likely to change as CGAT and the individual technologies on which it is built mature over time.

RESULTS

In overview, the CGAT tool works in two steps. In step one, CRISPR targets are identified in a user-specified sequence of interest with the sequence being pasted into a text field or selected from a list of gene/gene model names from the species of interest. In the second step, potential off-targets are identified. These two functionalities encompass the following steps:

1. For each genome available to search above, the genome sequence has been parsed in advance for valid CRISPR target sequences. All found target

sequences were exported to a SQL database along with some relevant metadata. Additionally, the transcript data for each gene has also been stored in the SQL DB for easy retrieval when a user opts to select the input sequence from a specific gene.

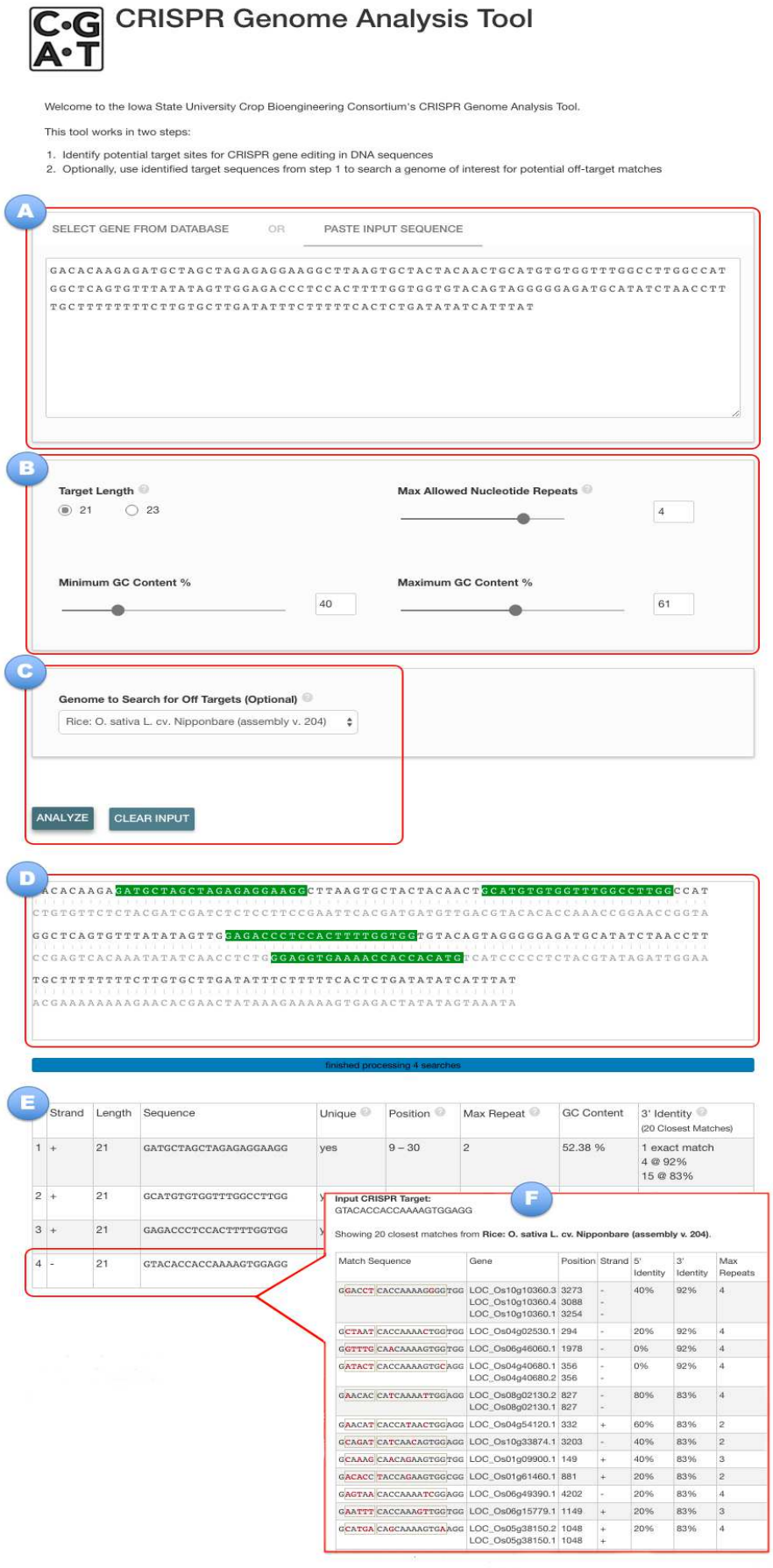
2. In the tool interface, Javascript is used to parse both the input sequence and its complement for valid CRISPR targets based on the user-provided search parameter (i.e., Target Length, GC Content and Allowed Nucleotide Repeats). The results are rendered in the browser and, for each found target sequence, a request is sent to the webserver to search the specified genome database for potential off-target matches.
3. For each request sent from the web browser to the webserver in the previous step, the server queries the database for the target genome with the user-provided search parameters.
4. Search results are filtered and sorted primarily by an identity score between an input subsequence (bases 6-18 for 21 base sequences or bases 6-20 for 23 base sequences) and the corresponding subsequences stored in the database. Additional sorting is performed based on an identity score between the subsequence at bases 2-5 of the input sequence and the corresponding subsequences in the database.
5. Finally, the webserver returns the search results to the browser, which updates the existing table. Clicking any table row reveals more details about the result.

OsSWEET11 EXAMPLE

The SWEET gene family of sugar transporters has been shown to play a vital role in multiple plant growth and developmental processes, including seed nutrition. They are also responsible for host recognition and subsequent sugar acquisition by the bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* - the causal agent of rice bacterial blight (Chen, 2014; Chen et al., 2010); (Boch et al., 2014). Jiang et al. demonstrated that efficient Cas9-mediated modification of the *OsSWEET11* promoter decreased pathogen-host interaction in rice (2013b). Here we search *japonica* rice (*Oryza sativa* L. cv. Nipponbare) for the same target as a representative usage example for CGAT.

As shown in Figure 2, the sequence for the *OsSWEET11* gene promoter (GenBank: CM000145.1 nucleotides 25503600-25503800) was used as input. CGAT default parameters are set to identify CRISPR targets of at least 21 nucleotides. The results table highlights potential CRISPR target regions in green. The *OsSWEET11* CRISPR target exploited by Jiang et al. (2013b) to induce mutation that increased host resistance to bacterial blight is the last in the group (i.e., sequence 5'-GTACACCACCAAAAGTGGAGG-3'). Next the targets are used to query for off-target matches genome-wide. No off-target 100% identical to the Jiang et al. target is identified in the rice genome.

FIGURE 2 (next page): CGAT example functionality using *OsSWEET11*. (A) Paste into the box a sequence (or select a sequence from the database). (B) Specify design parameters including target length, the maximum number of tandemly repeated nucleotides, and minimum/maximum GC content. (C) Select a genome to query for potential off- target recognition and hit the 'Analyze' button. (D) Visualize targets within the input sequence (displayed in green). (E) Evaluate and (F) prioritize targets using sequence identity as well as off-target sequence identity.


CRISPR Genome Analysis Tool
Welcome to the Iowa State University Crop Bioengineering Consortium's CRISPR Genome Analysis Tool.
This tool works in two steps:
1. Identify potential target sites for CRISPR gene editing in DNA sequences
2. Optionally, use identified target sequences from step 1 to search a genome of interest for potential off-target matches
A
SELECT GENE FROM DATABASE
OR
PASTE INPUT SEQUENCE
GACACAAGAGATGCTAGCTAGAGAGGAAGGCTTAAGTGCTACTACAACTGCATGTGTGGTTTGGCCTTGGCCAT
GGCTCAGTGTTTATATAGTTGGAGACCCTCCACTTTTGGTGGTGTACAGTAGGGGGAGATGCATATCTAACCTT
TGCTTTTTTTTTCTTGTGCTTGATATTTCTTTTTCACTCTGATATATCATTTAT
B
Target Length
21
23
Max Allowed Nucleotide Repeats
4
Minimum GC Content %
40
Maximum GC Content %
61
C
Genome to Search for Off Targets (Optional)
Rice: O. sativa L. cv. Nipponbare (assembly v. 204)
ANALYZE
CLEAR INPUT
D
finished processing 4 searches
E
Strand
Length
Sequence
Unique
Position
Max Repeat
GC Content
3' Identity
(20 Closest Matches)
1 + 21 GATGCTAGCTAGAGAGGAAGG yes 9 – 30 2 52.38 % 1 exact match 4 @ 92% 15 @ 83%
2 + 21 GCATGTGTGGTTTGGCCTTGG
3 + 21 GAGACCCTCCACTTTTGGTGG
4 - 21 GTACACCACCAAAAGTGGAGG
F
Input CRISPR Target:
GTACACCACCAAAAGTGGAGG
Showing 20 closest matches from Rice: O. sativa L. cv. Nipponbare (assembly v. 204).
Match Sequence
Gene
Position
Strand
5' Identity
3' Identity
Max Repeats
GGACCTCACCAAAAGGGTGG LOC_Os10g10360.3 3273 - 40% 92% 4
LOC_Os10g10360.4 3088 -
LOC_Os10g10360.1 3254 -
GCTAATCACCAAAACTGGTGG LOC_Os04g02530.1 294 - 20% 92% 4
GGTTTGCAACAAAAGTGGTGG LOC_Os06g46060.1 1978 - 0% 92% 4
GATACTCACCAAAAGTGCAGG LOC_Os04g40680.1 356 - 0% 92% 4
LOC_Os04g40680.2 356 -
GAACACCATCAAAATTGGAGG LOC_Os08g02130.2 827 - 80% 83% 4
LOC_Os08g02130.1 827 -
GAACATCACCATAACTGGAGG LOC_Os04g54120.1 332 + 60% 83% 2
GCAGATCATCAACAGTGGAGG LOC_Os10g33874.1 3203 - 40% 83% 2
GCAAAGCAACAGAAGTGGTGG LOC_Os01g09900.1 149 + 40% 83% 3
GACACCTACCAGAAGTGGCGG LOC_Os01g61460.1 881 + 20% 83% 2
GAGTAACACCAAAATCGGAGG LOC_Os06g49390.1 4202 - 20% 83% 4
GAATTTCACCAAAGTTGGTGG LOC_Os06g15779.1 1149 + 20% 83% 3
GCATGACAGCAAAAGTGAAGG LOC_Os05g38150.2 1048 + 20% 83% 4
LOC_Os05g38150.1 1048 +

CONCLUSIONS AND FUTURE WORK

The potential uses of CRISPRs for genome editing seem limitless. Their general utility across diverse species has sparked the rapid production of bioinformatics tools to predict and analyze target sequences across a multitude of genomes. In this review, we compared functionality among a list of CRISPR prediction software and described in detail how to use CGAT.

To enable generalized bioinformatics support of the CRISPR-Cas9 system across all species, emerging CRISPR analysis tools should be adapted to search public databases such as Genbank at NCBI directly. Other desirable features include reporting the presence of restriction enzyme cut sites in the target that may enable screening transformants by restriction digest of PCR products. Additionally reporting whether off-target matches represent duplicate genes and/or gene family members would be a useful feature.

ACKNOWLEDGEMENTS

We thank Darwin Campbell for help with computer administration and CGAT icon creation. Efforts by SZ were supported by the Iowa State University's Crop Bioengineering Consortium, a Presidential Initiative (described at http://cropbioengineering.iastate.edu). VAB was partially supported by the Iowa State University Graduate Minority Assistantship Program.

AUTHOR CONTRIBUTIONS

VAB: Developed design criteria and usage examples. Contributed heavily to writing the manuscript.

SZ: Developed design criteria, coded the CGAT tool, and contributed to writing the manuscript.

DW: Advised CGAT design and contributed to writing the manuscript.

YW, JL, and KC: Created a working CGAT proof-of-concept and approved the manuscript.

BY: Conceived of the tool, contributed to CGAT design, and contributed to writing the manuscript.

CJLD: Guided CGAT development and contributed heavily to writing the manuscript.

REFERENCES:

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 2007;*315*(5819), 1709–12. http://doi.org/10.1126/science.1138140

J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Annual Review of Phytopathology,* 2010;48 419-36.

Boch J, Bonas U, Lahaye T. TAL effectors--pathogen strategies and plant resistance engineering. *New Phytol,* 2014;204, 823-32

Boch J, Scholze H, Schornack S ... Bonas U. Breaking the code of DNA binding specificity of TAL-type III effectors. Science, 2009;326(5959)1509–12.

Bogdanove, AJ. Principles and applications of TAL effectors for plant physiology and metabolism. *Current Opinion in Plant Biology*, 2014;*19*, 99–104. http://doi.org/10.1016/j.pbi.2014.05.007

Carte J, Christopher RT, Smith JT, Olson S, Barrangou R, Moineau S, … Terns MP. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus. *Molecular Microbiology*, 2014;*93*(1), 98-112. http://doi.org/10.1111/mmi.12644

Chen LQ, Hou BH, Lalonde S, Takanaga H, Hartung ML, Qu X-Q, … Frommer WB. Sugar transporters for intercellular exchange and nutrition of pathogens. *Natur*e, 2010. 468(7323) 527–32. http://doi.org/10.1038/nature09606

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas, D. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 2010;186:757-61. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G. COSMID: A Web-

based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Molecular Therapy. Nucleic Acids*, 2014;*3*, e214. http://doi.org/10.1038/mtna.2014.64

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel, J, Charpentier E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011;471(7340), 602–607. http://doi.org/10.1038/nature09886

Gaj, T., Gersbach, C. A., Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. (2013). *Trends in Biotechnology*, *31*(7), 397–405. http://doi.org/10.1016/j.tibtech.2013.04.004

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, 2014; 196(4), 961–71. http://doi.org/10.1534/genetics.113.160713

Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 2007;35(Web Server issue), W52–7. http://doi.org/10.1093/nar/gkm360

Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nature Methods*, 2014;11(2), 122–3. http://doi.org/10.1038/nmeth.2812

Holkers M, Maggio I, Liu J, Janssen JM, Miselli F, Mussolino C, Recchia A, Cathomen T, Gonçalves MAFV. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Research*, 2013;41(5). http://doi.org/10.1093/nar/gks1446

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nature Biotechnology, 2013;31(3), 233–9. http://doi.org/10.1038/nbt.2508

Jiang W, Zhou H, Bi H, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Research, 2013;41(20), e188. http://doi.org/10.1093/nar/gkt780

Kaur K, Tandon H, Gupta AK, Kumar M. CrisprGE: a central hub of CRISPR/Cas-based genome editing. Database : The Journal of Biological Databases and Curation, 2015. http://doi.org/10.1093/database/bav055

Klug A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annu Rev Biochem 2010;79:213-31. doi: 10.1146/annurev- biochem-010909-095056

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science (New York, N.Y.), 1989;245(4918), 635–637. http://doi.org/10.1126/science.2503871

Lei Y, Lu L, Liu H-Y, Li S, Xing F, Chen LL. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. Molecular Plant, 2014;7(9), 1494–6. http://doi.org/10.1093/mp/ssu044

Li T, Huang S, Jiang WZ, Wright D, Spalding MH, Weeks DP, Yang B. TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucl. Acids Res., 2011;39:359-72.

Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual Review of Biochemistry, 2010;79, 181–211. http://doi.org/10.1146/annurev.biochem.052308.093131

Ma M, Ye AY, Zheng W, Kong L. A Guide RNA Sequence Design Platform for the CRISPR/Cas9 system for model organism genomes. BioMed Research International, 2013. http://doi.org/10.1155/2013/270805

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, … Koonin EV.Evolution and classification of the CRISPR-Cas systems. Nature Reviews. Microbiology, 2011;9(6), 467–77. http://doi.org/10.1038/nrmicro2577

Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Research, 2014;42(W1). http://doi.org/10.1093/nar/gku410

Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. Science 2009;326 (5959):1501.

Muller, H. J. Artificial transmutation of the gene. Science (New York, N.Y.), 1927;66(1699),84–87. http://doi.org/10.1126/science.66.1699.84

Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics (Oxford, England), 2014;31(7), 1120–3. http://doi.org/10.1093/bioinformatics/btu743

O'Brien A, Bailey TL. GT-Scan: identifying unique genomic targets. Bioinformatics (Oxford, England), 2014;30(18), 2673–5. http://doi.org/10.1093/bioinformatics/btu354

Peng Y, Clark KJ, Campbell JM, Panetta MR, Guo Y, Ekker SC. Making designer mutants in model organisms. Development, 2014;141(21), 4042–4054. http://doi.org/10.1242/dev.102186

Prykhozhij SV, Rajan V, Gaston D, Berman JN. CRISPR MultiTargeter: A Web Tool to Find Common and Unique CRISPR Single Guide RNA Targets in a Set of Similar Sequences. PloS One, 2015;10(3), e0119372. http://doi.org/10.1371/journal.pone.0119372

Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie, 2015 http://doi.org/10.1016/j.biochi.2015.03.025

Ruiz i Altaba A, Perry-O'Keefe H, Melton DA. Xfin: an embryonic gene encoding a multifingered protein in Xenopus. The EMBO Journal, 1987;6(10), 3065–70.

Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology, 2014;32(4), 347–55. http://doi.org/10.1038/nbt.2842

Singh R, Kuscu C, Quinlan A, Qi Y, Adli M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. Nucleic Acids Research, 2015. http://doi.org/10.1093/nar/gkv575

Stadler LJ. Mutations in barley induced by X rays and radium. Science, 1928;68,186–7

Stadler LJ. The Effect of X-Rays upon Dominant Mutation in Maize. Proceedings of the National Academy of Sciences of the United States of America, 1944;30(6), 123–8.

Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PloS One, 2015;10(4), e0124633. http://doi.org/10.1371/journal.pone.0124633

Upadhyay SK, Sharma S. SSFinder: High throughput CRISPR-Cas target sites prediction tool (2014). BioMed Research International, 2014. http://doi.org/10.1155/2014/742482

White FF, Potnis N, Jones JB, Koebnik R, 2009. The type III effectors of Xanthomonas. Plant Pathology, 2009;10(6)749-66

Wright DA, LI T, Yang B, Spalding MH. TALEN-mediated genome editing: prospects and perspectives (2014). Biochemical Journal, 2014;462(1), 15–24.

Xie S, Shen B, Zhang C, Huang X, Zhang Y. sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. PLoS ONE, 2014;9(6), e100448. http://doi.org/10.1371/journal.pone.0100448

Zhang Y, Ge X, Yang F, Zhang L, Zheng J, Tan X, Jin Z-B, Qu J, Gu F. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Scientific Reports, 2014;4, 5405. http://doi.org/10.1038/srep05405

Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: a bioconductor package

CHAPTER III

GENOME ENGINEERING TO REDUCE PEANUT ALLERGENICITY

ABSTRACT

Researchers estimate that up to fifteen million Americans have food allergies. Of those affected, peanut allergies account for the highest number of deaths per year of any other food borne allergen. Although several peanut proteins have been identified for their role in triggering human immune response, there are four main proteins responsible for the hypersensitive reaction: Ara h 1, Ara h 2, Ara h 3, and Ara h 6. Anaphylaxis (an acute allergic reaction) is triggered by peanut in sensitive individuals, by hypersensitive immune response to the Ara h 2 protein, which has been found to be the most severely allergenic of the 4 genes. Here we use the CRISPR Genome Analysis Tool (CGAT) along with other bioinformatics techniques to propose a methodology for the targeted deletion of the Ara h 1 peanut allergen.

INTRODUCTION

Many members of the Leguminisae family (also called legumes or beans) of flowering plants are capable of forming symbiotic relationships with the nitrogen-fixing bacteria *rhizobium* (Phillips, 1980). Grain legumes are some of the richest sources of plant-based protein in the world (USDA Nutrient Database). They also serve as an abundant source of oil and micronutrients. Crops within this family include common bean, soybean, chickpea, lentil, and peanut.

The National Peanut Board estimates that global peanut production totals about 29 million metric tons per year. The United States is a leading peanut producer – third in the world following China and India (http://www.nationalpeanutboard.org). Studies indicate that peanut is

a rich source of monounsaturated fats and other nutrients, such as niacin and manganese, which have been linked to a reduced risk of heart related complications (Kris-Etherton et al., 2008). Aside from increased heart health, studies indicate that peanuts can be beneficial in preventing colon cancer as well as gallstones. Data published in the journal of Neurology, Neurosurgery, and Psychiatry suggest that peanut consumption may also play an important role in protecting individuals from the neurodegenerative properties of Alzheimer's and other diseases.

Given the relatively low cost of peanut and its high availability for consumption, it is unfortunate that up to fifteen million Americans have food allergies. Of those affected, peanut allergies account for the highest number of deaths per year of any other food borne allergen (Finkelman, 2010). Indeed, peanut is becoming increasingly prevalent in industrial food processing (Chang et al., 2013; Zhao et al., 2012), making it increasingly difficult for families to provide healthy and safe foods for affected children and adults alike.

In allergic individuals, the immune system responds to harmless foreign molecules causing a range of allergy symptoms from dermatitis to anaphylaxis and even death (Finkelman, 2010). Upon ingestion, peanut allergens are introduced to the immune system via the mucosal lining of the abdomen. Epithelial cells, along the interior wall of the abdomen, transfer allergens to dendritic cells for further processing. Dendritic cells initiate degradation of peanut allergens into smaller peptide fragments that are exposed on the cell surface via the major histocompatibility complex (MHC) (He et al., 2015). Naïve T-cell receptors recognize the MHC-peptide complex, and mediate the activation of specialized T-cells to trigger immune response (known as T-cell priming). T-cell priming prompts the release of interleukins IL-4, IL-5, and IL-13. Cytokine signaling encourages differentiation of B cells that mediate the production of allergen specific Immunoglobulin E (IgE) antibodies. Peanut specific IgE

antibodies are tightly bound to the surface of mast cells, and act as extracellular allergen receptors (He et al., 2015). Subsequent exposure to peanut protein triggers the release of anti-inflamatory molecules associated with the tell-tale symptoms (i.e., swelling of the extremities, asthma, diarrhea, vomiting, and anaphylaxis; He et al. 2015; Finkelman 2010).

Allergenicity is described as a measure of the ability to elicit an immune response. A total of eleven proteins are potentially involved in peanut allergenicity, four of which have been identified as the most important based on clinical tests - immunoblots, skin prick tests, *ex viva* basophil histamine release assays. They are Ara h 1, Ara h 2, Ara h 3 and Ara h 6 (Koppelman et. al. 2010). Ara h 1 is one of the most extensively studied proteins of the Ara allergen quartet. It affects 35-95% of allergic individuals. Ara h 1 is a 65-kDa vicilin, which has a homo-trimeric structure (Shin et al., 1998). Like other seed storage proteins Ara h 1 comprises up to 12 -15% of the total peanut protein content. The concentration of Ara h 1 in peanuts increases with kernel size and protein expression is linked with peanut maturity. There are 21 linear epitopes on mature protein monomers with 14 in the core region of the trimer. The core region is buried during trimer formation, which could potentially explain its reduced activity relative to other major allergens (Cabanos et al., 2011; Chruszcz et al., 2011; Zhou et al., 2013).

Ara h 3 is a 60 kDA trypsin inhibitor in the glycinin family. Ara h 3 is responsible for seed longevity and deterring the catabolic activity of proteases present in the abdomen of most predators (Koppelman et al., 2003; Zhou et al., 2013).

Ara h 2 and Ara h 6 are seed storage proteins in the conglutin family. Clinical studies suggest that the 2S albumin Ara h 2 is the most severely allergenic of the two because of its ability to trigger inflammatory response from human basophils in greater than 90% of peanut allergic individuals (Maleki et. al. 2013). Ara h 2 has two isoforms (Ara h 2.01 and Ara h 2.02)

and a 59% amino acid homology with Ara h 6. Both Ara h 2 and Ara h 6 have multiple disulphide-bridged cysteine residues that result in a tightly coiled, protease resistant, and heat stable core that plays a role in allergenicity. These antigens are particularly dangerous due to their ability to form intermolecular cross-links when roasted, making the allergic reaction more severe when people with peanut allergies ingest roasted peanuts. (Chu et. al. 2008) (Chen et. al. 2013) (Li et. al. 2013) (Klemans et. al. 2013)

Cultivated peanut (*Arachis hypogea*) is an allotetraploid (2x = 2n =40) in the genus *Arachis*, consisting of approximately 69 species spanning 9 morphological classes, which originated in South America. Karyotyping and chromosomal morphology data presented by Krapovickas et al. (1994) concluded that *Arachis hypogea* likely arose from diploid progenitors *Arachis duranensis* (genome symbol AA) and *Arachis ipeansis* (genome symbol BB) (Krapovickas & Gregory 1994; Jung 2003). The major antigens in peanut (*Arachis hypogea*) are encoded by both the A and B genomes.

(Dodo et al., 2008) demonstrated the successful application of RNA interference (RNAi) technology targeting the Ara h 2 peanut protein. Results indicated an overall reduction of Ara h 2 content in crude peanut extract, as well as a significant reduction of IgE-Arah2 recognition during immunoblot assays. Although these findings demonstrated the feasibility of RNAi to mediate a reduction in allergenicity in peanut, RNAi has multiple drawbacks. For instance, siRNAs generally result in reduced gene expression or “knockdown” at the transcript level (as opposed to an absence of the presence of the antigens), and are associated with a wide range of transient phenotypic variance (Senthil-Kumar and Mysore, 2011).

In these studies we associate the known peanut allergens with their genomic location in the *Arachis hypogea* subgenomes in an effort to determine whether the allergens could be

targeted for gene editing. For Ara h 1, we design a CRISPR-Cas9 construct for future experimentation and to serve as a test case toward developing reduced allergenicity peanut cultivars.

METHODS

Genomic location of known peanut allergens

Allergen sequence information for all known *Arachis* proteins were accessed via the Allergen.org database. Genomic locations of Ara proteins relative to each *Arachis* pseudomolecule were mapped using protein input sequences and BLASTp functionality (Altschul, 2005) available via PeanutBase.org (Dash et al 2016). A protein-protein alignment was generated for each query against known proteins in each progenitor genome using default parameters.

Identification of CRISPR targets using CGAT

Subcloning and sequence analysis of the Ara h 1 gene was previously described by Viquez et al. (2004). CRISPR targets were identified using the CRISPR Genome Analysis Tool (CGAT; described in Chapter 2). CGAT parameters were adjusted to assess 21 nucleotide targets with a maximum nucleotide repetition of 3, and overall GC content between 45 and 65%. Guide RNAs exhibiting single base pair mismatches within the eleven nucleotides adjacent to the 3' PAM sequence have been shown to abolish Cas9 recognition (Cong et al. 2013). For this reason, potential for constructs to create off-target effects were evaluated based on percentage non-target 3' identity to the genuine target sequence.

RESULTS

Table 1 lists the genomic location of each of the known peanut allergens. Here we focus on the major allergens Ara h 1, Ara h2, Ara h3, and Ara h6. Ara h 1 (Genbank accession ADQ53858.1) has homeologs on chromosome nine of the A and B genomes associated with positions A09:111191611 - 111193663 and B09:145805360 – 145807488 in the VERSION peanut assembly. Although isoforms have been identified for Ara h 2 in in the peanut genome, no clear homologs could be identified in the current *Arachis* sequence assemblies. Ara h 3 (Genbank accession AAC63045) has three tandem homeologs located on chromosome 6 of the A and B genomes. These are associated with positions A06:1,263,038 - 1,265,555, A06:1,278,069 - 1,286,732, A06:1,778,238 - 1,780,113, B06:21,966,364 - 21,968,437, B06:22,154,349 - 22,156,660, B06:22,165,994 - 2,172,682. Ara h 6 (Genbank accession AAD56337) has homeologs on chromosome eight of the A and B genomes associated with positions A08:34,023,496 - 34,113,800 and B08:14,077,897 - 14,223,923, respectively.

Of the major allergens, Ara h 1 was selected for further analysis due to its performance in clinical trials to elicit immune response, as well as its localization to a single chromosomal location on both progenitor genomes at locations A09:111191611 - 111193663 and B09:145805360 - 145807488 as mentioned above.

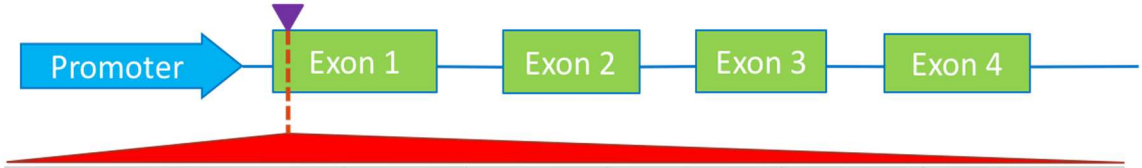
TABLE 1. Location of the major peanut allergen homeologs

ALLERGEN NAMES	GENBANK	FUNCTION	kDA	Patient Recognition	LOCUS	HOMEOLOGS
Ara H1	ADQ53858.1	provide nourishment for seedlings; vicilin	63.5	65%	A09:111191611..111193663 B09:145805360..145807488	one from the A and B genomes
Ara H2	AAK96887 - 2.01 AAN77576.1 - 2.02	conglutin, protease inhibitor/seed storage/LTP family; defense against pathogens	16.6	85%	none found	one from the A and B genomes. Some reference to it originating in the B genome
Ara H3/4	AAC63045	11S seed storage protein in the glycinin family; provide nourishment for seedlings	60	53%	A06:1,263,038..1,265,555, A06:1,278,069..1,286,732, A06:1,778,238..1,780,113, B06:21,966,364..21,968,437, B06:22,154,349..22,156,660, B06:22,165,994..22,172,682	three copies from A genome and three copies from B genome
Ara H5	AAD55587	prolifin; regulates polymerization of actin monomers;	14	13%	A05:89160862..89168923 B05:139588557..139597435	one from the A and B genomes
Ara H6	AAD56337	defense against pathogens; source of amino acids for seedling growth	15	38%	A08:34,023,496..34,113,800 B08:14,077,897..14,223,923	two copies in the B genome and one copy in the A genome
Ara H7	AAD56719.1	conglutin 2S storage protein	16.4	43%	A10:103,211,665..103,211,797 B10:129,752,017..129,752,149	one from the A and B genomes
Ara H8	AAQ91847.1	possible delivery vehicle for flavonoids; pathogen defense	17	NA	A01:96132311..96133187 B01:130794216..130795263	one from the A and B genomes
Ara H9	ABX56711	defense against pathogens; helps create hydrophobic layers in plant;Non-specific lipid-transfer protein	9.8	NA	A02:84440368..84441960 B02:96749005..96751079	one from the A and B genomes
Ara H10	AAU21499	structural proteins of oil bodies in seeds; oleosin0	16	NA	none found	
Ara H11	AAZ20276/7	structural proteins of oil bodies in seeds; oleosin1	14	NA	none found	
(Kleber-Janke et al. 1999)						

Viquez et al. (2003) reported the sequence analysis and genomic structure of an Ara h 1 clone. Results indicated an open reading frame containing four exons and three introns. The 5' UTR directly flanking the first exon is a 1,926 bp promoter, harboring 17 cis regulatory elements. Of the elements described, the strength of the Ara h 1 promoter is suggested by the existence of multiple pairs of transcription factor binding sites. Because of the great deal of functional overlap that exists between pairs of regulatory elements in the Ara h 1 promoter (multiple CAAT boxes, multiple G boxes, and multiple enhancers), targeted disruption of a single element would not be likely to knock out expression of the Ara h 1 gene. With this in mind, CGAT sgRNA

prediction was targeted to the 5' end of the first exon of each of two peanut homeologs (i.e., the *A. duranensis* and *A. ipaensis* copies) of the Ara h 1 gene (see Figures 2 and 3). Sequence analysis of this Ara h 1 clone against the *A. duranensis* and *A. ipaensis* genomes indicated a 100% identity match to genes Adur.2H0R0 and Araip.T8528 in *A. duranensis* and *A. ipaensis* genomes, respectively.

A.



B.

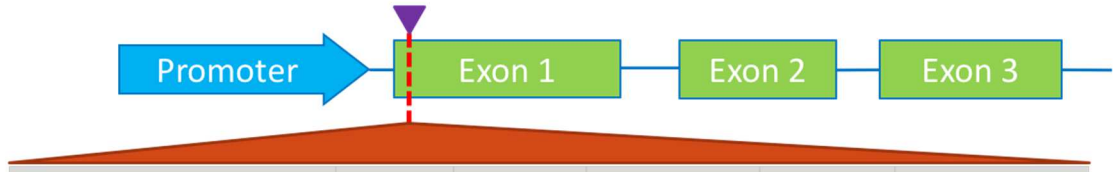
sgRNA sequence	Unique?	Position	Max Repeats	GC %	Off-targets
GTAGTAGCAGGAGCAATGAGAGG	yes	47 – 70	2	52.17 %	1 exact match 10 @ 86% 9 @ 79%

C.

	Match Sequence	Gene	Position	Strand	5' Identity	3' Identity	Max Repeats
A.duranensis	GTAGTA GCAGGAGCAATGAGAGG	Aradu.2H0R0+	47	+	100%	100%	2
	GAGGTA GCAGCAGCAACGAGTGG	Aradu.5PD0E+	5008	-	60%	86%	2
A.ipaensis	Match Sequence	Gene	Position	Strand	5' Identity	3' Identity	Max Repeats
	GTAGTA GCAGGAGCAATGAGAGG	Araip.T82B5-	2099	-	100%	100%	2
	GTAGTT GCAGGAGCAATAACGGG	Araip.T7PRK-	451	-	80%	86%	3

FIGURE. 2: Targeting Ara h 1 using the CGAT design tool. (A): Ara h 1 gene model in *A. duranensis* progenitor genome. The Ara h 1 gene model is composed of a promoter (blue), four exons (green boxes), and three introns (blue line between exons). (B) A single sgRNA (5'- GTAGTAGCAGGAGCAATGAGAGG - 3') corresponding to the genomic location +47 on exon 1 was selected for further analysis (approximate position denoted by a red dotted line). (C) Off-target analysis of the sgRNA (5'- GTAGTAGCAGGAGCAATGAGAGG - 3') shows that the only identical targets are the Ara h 1 homeologs (i.e., Aradu.2H0R0+ and Araip.T82B5-).

A.



B.

sgRNA sequence	Unique?	Position	Max Repeats	GC %	Off-targets
GTAGTAGCAGGAGCAATGAGAGG	yes	2099 – 2122	2	52.17 %	1 exact match 12 @ 86% 7 @ 79%

C.

A. ipaensis

Match Sequence	Gene	Position	Strand	5' Identity	3' Identity	Max Repeats
GTAGTA GCAGGAGCAATGAG AGG	Araip.T82B5-	2099	-	100%	100%	2
GTAGTT GCAGGAGCAATAAC GGG	Araip.T7PRK-	451	-	80%	86%	3

A. duranensis

Match Sequence	Gene	Position	Strand	5' Identity	3' Identity	Max Repeats
GTAGTA GCAGGAGCAATGAG AGG	Aradu.2H0R0+	47	+	100%	100%	2
GAGGTA GCAGCAGCAACGAG TGG	Aradu.5PD0E+	5008	-	60%	86%	2

FIGURE. 3: Targeting Ara h 1 using the CGAT design tool
(A): Ara h 1 gene model in *A. ipaennsis* progenitor genome. Ara h 1 gene model composed of a promoter (blue), four exons (green boxes) and three introns (blue line between exons). (B) A single sgRNA (5'- GTAGTAGCAGGAGCAATGAGAGG - 3') corresponding to the genomic location -2099 on exon 1 was selected for further analysis (approximate position denoted by a red dotted line). (C) (5'- GTAGTAGCAGGAGCAATGAGAGG - 3') shows that the only identical targets are the Ara h 1 homeologs (i.e., Aradu.2H0R0+ and Araip.T82B5-).

Separate sgRNA predictions were performed for Ara h 1 homeologs Adur.2H0R0 and Araip.T82B5 against both the A and B progenitor genomes. Multiple targets were evaluated, and a single guide RNA capable of targeting the proximal region of the first exon in both Ara h 1 homeologs was selected for further analysis.

Sample sgRNA constructs would be designed following the arrangement shown in Figure 4. sgRNA gene expression is driven by the U6 promoter. The protospacer adjacent motif (dark green) allows the Cas9 endonuclease to target and create doublestrand breaks at a genomic loci corresponding to the guide RNA in question. The strong CMV (Cauliflower Mosaic Virus) promoter is used downstream of the guide RNA to drive expression of the Cas9 endonuclease. GFP (green fluorescent protein) can be attached to the end of the cassette for quick identification of positive transformants after cloning.

CONCLUSIONS AND FUTURE WORK

Peanut is an agriculturally important species. Unfortunately, the prevalence of peanut allergy continues to increase across multiple countries – particularly amongst school-aged children(). Despite national guidelines for managing school related incidence of allergen exposure, the increased use of peanut and peanut extracts in food processing continues to present a growing public health concern – especially for parents of affected children. School districts across the U.S. have taken proactive measures to enforce strict bans of nuts and nut products on school premises. While this presents a win for concerned parents, many experts agree that this sort of solution will not completely protect children with extreme allergies.

By designing constructs that aim to knock out the specific allergens in peanut, we hypothesize a possible alternative to complete avoidance of situations where peanuts are present is given. While it is not reasonable to recommend that children allergic to peanut eat peanuts that have the major allergens knocked out, the potential for a classmate inadvertently exposing an allergic child to an allergen by way of e.g., eating a peanut butter sandwich nearby is reduced, making the lunch room a safer place.

Gene editing of cultivated peanut has several obstacles including overall genomic complexity and involvement of many epitopes in the allergenicity complex. Multiple techniques have been applied to reducing peanut allergenicity including enzymatic exposure and RNA interference. Here we aimed to completely knock out the major allergen Ara h 1 by focusing on constructs that target the 5' end of copies of the gene in both the A and B genomes. Using CGAT and other bioinformatics analyses, a single sgRNA was selected for further analysis using experimental molecular techniques. Collaborators Peggy Ozias-Akins and Don Weeks at the University of Georgia and University of Nebraska, respectively, will receive these predictions. This would enable them to design guide RNA constructs targeting the Ara h 1 gene for further analysis (Fig 4).

Example Cas9 construct harboring sgRNA of interest

FIGURE 4. Possible Cas9 sgRNA construct. U6 promoter (red) drives expression of the sgRNA (yellow). CMV promoter (purple) is used to drive expression of Cas9 endonuclease (blue). GFP (light green) used as a reporter gene to easly identify positive transformants.

Next steps for this work involve the creation of sgRNA constructs, introduction of constructs into peanut cultivars, and testing to insure that the genes of interest have been knocked out. Upon introduction of this Cas9 construct, double strand breaks are predicted to occur at nucleotide positions 29 – 52 of the first exon of the two Ara h 1 homeologs (in both the A and B genomes). Recovery of repaired double strand breaks would require error prone non-homologous end joining. Nucleotide changes that result in a large deletion or a premature stop codon would be

ideal for downstream analyses. Downstream analyses would aim to elucidate whether functional protein knockouts produce viable seeds, as well as to determine whether flavor, mouth feel, nutrition, and seed filling are adversely affected by editing the Ara h 1 gene (Essam et al. 1982). (Dienner et al. 1982).

CHAPTER IV

CONCLUSIONS

In this thesis I described our efforts to create a generalized tool for gene editing then applied it (and other bioinformatics tools) to the problem of knocking out the major peanut allergen Ara h 1. "Hypoallergenicity" refers to a reduction in potential for an allergic reaction compared to other products (http://fda.gov). Multiple processes have been described to reduce human immuno-sensitivity to peanut proteins. One strategy is to select peanut cultivars with decreased allergen presence and introduce them into conventional breeding programs to decrease the total amount of antigen present in commercial peanut. Other strategies involve exposing peanut to enzymes or polymers that greatly reduce the presence of consequential proteins (Finkelman, 2010; Yu et al., 2011). In a practical sense, the long-term goal of Chapter 3 is not to create a *hypoallergenic* peanut, per se. Because there are many proteins involved and individuals' sensitivity to different epitopes varies, even a peanut that has the major allergens knocked out cannot be guaranteed 'hypoallergenic' for any and all individuals. Instead, the two main goals of the work described in Chapter 3 are (1) to design a peanut that has *reduced* allergenicity in order to create cultivars that could be used to increase food safety and (2) to demonstrate a clearly beneficial use of biotech that is a consumer-centric win.

Both Chapters 2 and 3 relevant to a major discussion ongoing in public forums regarding the safety of biotech food. A recent study suggests that the overt public opposition to biotech food may be due to three major factors: (1) psychological essentialism, (2) misplaced emotion, and (3) teleological thinking (Blancke et al., 2015). Because none of these issues of perception

can readily be fully addressed using only facts, figures, and experimental data, it is clear that changing public perception toward greater support of biotech in the marketplace will require the concerted efforts of educators, extension specialists, outreach coordinators, and many others.

REFERENCES

Abelson, P.H. (1992). Biotechnology in a global economy. Office of Technology Assessment Report *255*, 381.

Altschul, S.F. (2005). BLAST Algorithm.

Beerli, R.R., and Barbas, C.F. (2002). Engineering polydactyl zinc-finger transcription factors. Nature Biotechnology *20*, 135–141.

Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J., and Van Montagu, M. (2015). Fatal attraction: the intuitive appeal of GMO opposition. Trends in Plant Science *20*, 414–418.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science (New York, N.Y.) *326*, 1509–1512.

Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors - pathogen strategies and plant resistance engineering. New Phytologist *204*, 823–832.

Bogdanove, A.J. (2014). Principles and applications of TAL effectors for plant physiology and metabolism. Current Opinion in Plant Biology *19*, 99–104.

Boudry, M., Blancke, S., and Pigliucci, M. (2014). What makes weird beliefs thrive? The epidemiology of pseudoscience. Philosophical Psychology 1–22.

Cabanos, C., Urabe, H., Tandang-Silvas, M.R., Utsumi, S., Mikami, B., and Maruyama, N. (2011). Crystal structure of the major peanut allergen Ara h 1. Molecular Immunology *49*, 115–123.

Carte, J., Christopher, R.T., Smith, J.T., Olson, S., Barrangou, R., Moineau, S., Glover, C.V.C., Graveley, B.R., Terns, R.M., and Terns, M.P. (2014). The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in Streptococcus thermophilus. Molecular Microbiology *93*, 98–112.

Chang, A.S., Sreedharan, A., and Schneider, K.R. (2013). Peanut and peanut products: A food safety perspective. Food Control *32*, 296–303.

Chen, L.-Q. (2014). SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytologist *201*, 1150–1155.

Chen, L.-Q., Hou, B.-H., Lalonde, S., Takanaga, H., Hartung, M.L., Qu, X.-Q., Guo, W.-J., Kim, J.-G., Underwood, W., Chaudhuri, B., et al. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. Nature *468*, 527–532.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector

nucleases. Genetics *186*, 757–761.

Chruszcz, M., Maleki, S.J., Majorek, K. a., Demas, M., Bublin, M., Solberg, R., Hurlburt, B.K., Ruan, S., Mattisohn, C.P., Breiteneder, H., et al. (2011). Structural and immunologic characterization of Ara h 1, a major peanut allergen. Journal of Biological Chemistry *286*, 39318–39327.

Ciftci, K. (2000). Applications of genetic engineering in veterinary medicine. Advanced Drug Delivery Reviews *43*, 57–64.

Cradick, T.J., Qiu, P., Lee, C.M., Fine, E.J., and Bao, G. (2014). COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. Molecular Therapy. Nucleic Acids *3*, e214.

Esam A. Ahmed and Clyde T. Young. Composition, nutrition, and flavor of peanuts. Ch 17 In Harold T. Pattee and Clyde T. Young (eds.), Peanut Science and Technology (1982). P 655-688. American Peanut Research and Education Society, Yoakum, TX.

Datta, A. (2013). Genetic engineering for improving quality and productivity of crops. Agriculture & Food Security *2*, 15.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Diener, Urban L., Robert E. Pettit, and Richard J. Cole. Aflatoxins and other mycotoxins in peanuts. Ch 13 In Harold T. Pattee and Clyde T. Young (eds.), Peanut Science and Technology (1982). P 486-519. American Peanut Research and Education Society, Yoakum, TX.

Dodo, H.W., Konan, K.N., Chen, F.C., Egnin, M., and Viquez, O.M. (2008). Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. Plant Biotechnology Journal *6*, 135–145.

Finkelman, F.D. (2010). Peanut allergy and anaphylaxis. Current Opinion in Immunology *22*, 783–788.

Gaj, T., Gersbach, C.A., and Barbas, C.F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology *31*, 397–405.

Gersbach, C. a, Phillips, J.E., and García, A.J. (2007). Genetic engineering for skeletal regenerative medicine. Annual Review of Biomedical Engineering *9*, 87–119.

Giddings, V. (2015). A Policymaker's Guide to the GMO Controversies. 19.

Glick, B.R., and Pasternak, J.J. (2003). Molecular biotechnology: principles and applications of recombinant DNA.

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., Donohue, L.K., Cummings, A.M., and O'Connor-Giles, K.M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics *196*, 961–971.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Research *35*, W52–W57.

He, J., Narayanan, S., Subramaniam, S., and Ho, W.Q. (2015). Biology of IgE Production: IgE Cell Differentiation and the Memory of IgE Responses. In IgE Antibodies: Generation and Function, J.J. Lafaille, and M.A. Curotto de Lafaille, eds. (Cham: Springer International Publishing), pp. 1–19.

Heigwer, F., Kerr, G., and Boutros, M. (2014). E-CRISP: fast CRISPR target site identification. Nature Methods *11*, 122–123.

Holkers, M., Maggio, I., Liu, J., Janssen, J.M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T., and Gonçalves, M.A.F. V (2013). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. Nucleic Acids Research *41*.

James, C. (2014). Executive summary. Global status of Commercialized Biotech/GM Crops : 2013.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D.P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. Nucleic Acids Research *41*, e188.

Jung, S. (2003). The Phylogenetic Relationship of Possible Progenitors of the Cultivated Peanut. Journal of Heredity *94*, 334–340.

Kaur, K., Tandon, H., Gupta, A.K., and Kumar, M. (2015). CrisprGE: a central hub of CRISPR/Cas-based genome editing. Database : The Journal of Biological Databases and Curation *2015*.

Klug, A. (2010). The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation.

Koppelman, S.J., Knol, E.F., Vlooswijk, R.A.A., Wensing, M., Knulst, A.C., Hefle, S.L., Gruppen, H., and Piersma, S. (2003). Peanut allergen Ara h 3: isolation from peanuts and biochemical characterization. Allergy *58*, 1144–1151.

KRAPOVICKAS, A., and GREGORY, W.C. (1994). TAXONOMIA DEL GENERO "ARACHIS (LEGUMINOSAE)." Bonplandia *8*, 1–186.

Kris-Etherton, P.M., Hu, F.B., Ros, E., and Sabaté, J. (2008). The role of tree nuts and peanuts in the prevention of coronary heart disease: multiple potential mechanisms. The Journal of Nutrition *138*, 1746S – 1751S.

Lee, M.S., Gippert, G.P., Soman, K. V, Case, D.A., and Wright, P.E. (1989). Three-dimensional solution structure of a single zinc finger DNA-binding domain. Science (New York, N.Y.) *245*, 635–637.

Lei, Y., Lu, L., Liu, H.-Y., Li, S., Xing, F., and Chen, L.-L. (2014). CRISPR-P: a web tool for

synthetic single-guide RNA design of CRISPR-system in plants. Molecular Plant *7*, 1494–1496.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Research *39*, 359–372.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual Review of Biochemistry *79*, 181–211.

Ma, M., Ye, A.Y., Zheng, W., and Kong, L. (2013). A Guide RNA Sequence Design Platform for the CRISPR/Cas9 system for model organism genomes. BioMed Research International *2013*.

Mak, A.N.-S., Bradley, P., Bogdanove, A.J., and Stoddard, B.L. (2013). TAL effectors: function, structure, engineering and applications. Current Opinion in Structural Biology *23*, 93–99.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nature Reviews. Microbiology *9*, 467–477.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Research *42*.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science (New York, N.Y.) *326*, 1501.

Muller, H.J. (1927). ARTIFICIAL TRANSMUTATION OF THE GENE. Science (New York, N.Y.) *66*, 84–87.

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2014). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics (Oxford, England) *31*, 1120–1123.

O'Brien, A., and Bailey, T.L. (2014). GT-Scan: identifying unique genomic targets. Bioinformatics (Oxford, England) *30*, 2673–2675.

Peng, Y., Clark, K.J., Campbell, J.M., Panetta, M.R., Guo, Y., and Ekker, S.C. (2014). Making designer mutants in model organisms. Development *141*, 4042–4054.

Phillips, D.A. (1980). Efficiency of Symbiotic Nitrogen Fixation in Legumes. Annual Review of Plant Physiology *31*, 29–49.

Phillips, P.W.B. (2002). Biotechnology in the global agri-food system. Trends in Biotechnology *20*, 376–381.

Prykhozhij, S. V, Rajan, V., Gaston, D., and Berman, J.N. (2015). CRISPR MultiTargeter: A Web Tool to Find Common and Unique CRISPR Single Guide RNA Targets in a Set of Similar Sequences. PloS One *10*, e0119372.

Radakovits, R., Jinkerson, R.E., Darzins, A., and Posewitz, M.C. (2010). Genetic engineering of algae for enhanced biofuel production. Eukaryotic Cell *9*, 486–501.

Ruiz i Altaba, A., Perry-O'Keefe, H., and Melton, D.A. (1987). Xfin: an embryonic gene

encoding a multifingered protein in Xenopus. The EMBO Journal *6*, 3065–3070.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology *32*, 347–355.

Senthil-Kumar, M., and Mysore, K.S. (2011). Caveat of RNAi in plants: the off-target effect. Methods in Molecular Biology (Clifton, N.J.) *744*, 13–25.

Shin, D.S., Compadre, C.M., Maleki, S.J., Kopper, R. a, Sampson, H., Huang, S.K., Burks, a W., and Bannon, G. a (1998). Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein. The Journal of Biological Chemistry *273*, 13753–13759.

Singh, R., Kuscu, C., Quinlan, A., Qi, Y., and Adli, M. (2015). Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. Nucleic Acids Research.

Stadler, L.J. (1928). Mutations in barley induced by X rays and radium. Science *68*, 186–187 ST – Mutations in barley induced by X ray.

Stadler, L.J. (1944). The Effect of X-Rays upon Dominant Mutation in Maize. Proceedings of the National Academy of Sciences of the United States of America *30*, 123–128.

Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L. (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PloS One *10*, e0124633.

Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. Nature Biotechnology *30*, 593–595.

Upadhyay, S.K., and Sharma, S. (2014). SSFinder: High throughput CRISPR-Cas target sites prediction tool. BioMed Research International *2014*.

Viquez, O.M., Konan, K.N., and Dodo, H.W. (2003). Structure and organization of the genomic clone of a major peanut allergen gene, Ara h 1. Molecular Immunology *40*, 565–571.

Wright, D.A. (Iowa S.U., LI, T. (Iowa S.U., Yang, B. (Iowa S.U., and Spalding, M.H. (Iowa S.U. (2014). TALEN-mediated genome editing: prospects and perspectives. Biochemical Journal *462*, 15–24.

Xie, S., Shen, B., Zhang, C., Huang, X., and Zhang, Y. (2014). sgRNAcas9: A Software Package for Designing CRISPR sgRNA and Evaluating Potential Off-Target Cleavage Sites. PLoS ONE *9*, e100448.

Yu, J., Ahmedna, M., Goktepe, I., Cheng, H., and Maleki, S. (2011). Enzymatic treatment of peanut kernels to reduce allergen levels. Food Chemistry *127*, 1014–1022.

Zhang, Y., Ge, X., Yang, F., Zhang, L., Zheng, J., Tan, X., Jin, Z.-B., Qu, J., and Gu, F. (2014). Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Scientific Reports *4*, 5405.

Zhao, X., Chen, J., and Du, F. (2012). Potential use of peanut by-products in food processing: A review. Journal of Food Science and Technology *49*, 521–529.

Zhou, Y., Wang, J., Yang, X., Lin, D., Gao, Y., Su, Y., Yang, S., Zhang, Y., and Zheng, J. (2013). Peanut Allergy , Allergen Composition , and Methods of Reducing Allergenicity : A Review. *2013*.

Zhu, L.J., Holmes, B.R., Aronin, N., and Brodsky, M.H. (2014). CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. PloS One *9*, e108424.