From: Scott Allen Jackson
To: Daniel O"Connor
Subject: Re: Peanut Enrichment Australia
Date: Thursday, October 02, 2014 1:36:04 PM

Hi Dan, what kind of help are you looking for? DO you want to see if they cover genes? You may want to talk to Steven Cannon.
scott

On Oct 1, 2014, at 8:26 PM, Daniel O'Connor <d.oconnor@uq.edu.au> wrote:

> Hi Scott,
> I have been working with Nimblegen to design probes for enrichment of Arachis hypogaea.
> Initially I sent them chromosome 1 of both diploids to get started and they have supplied a basic design based on these, see attached and email below.
> Just wondering who would be the best person in the consortium to maybe discuss these with as I have very limited experience with these and would like to make sure I am on the right track.
> I have contacted Rajeev Varshey, as you mentioned below, and am waiting for a reply.
> Thanks for your time.
> Regards,
> Dan
>
> **From:** Scott Allen Jackson [mailto:sjackson@uga.edu]
> **Sent:** Tuesday, 22 July 2014 11:15 AM
> **To:** Daniel O'Connor
> **Cc:** Steven Cannon
> **Subject:** Re: peanut assemblies
> **Importance:** High
>
> Hi Dan, Publication should wait until the gneome is published. Hopefully that is 2015:) also, you may want to check with Rajeev Varshney as I think they too were developing some enrichment tools, possibly with Roche.
>
> scott
>
> On Jul 21, 2014, at 7:04 PM, Daniel O'Connor <d.oconnor@uq.edu.au> wrote:
>
> > Hi Steven & Scott,
> >
> > I am working towards getting the enrichment of my poor and excellent blanching accessions organised in the near future.
> >
> > I would like make sure that I have permission to supply data from the progenitor preliminary gene models to Roche Nimblegen for the design of the enrichment probes.

There will be no publications on any of this until 2015. I understand that I will need to work with the PGC to ensure all the correct protocol publications are followed.

Thanks very much for your time.

Regards

Dan O'Connor

Research Scholar
The University of Queensland
Queensland Alliance for Agriculture and Food Innovation (QAAFI)
J Bjelke-Peteresen Research Station, Goodger Rd, Kingaroy, 4610
Tel: +61 7 4160 0723 Fax: +61 7 4162 3238
Mobile: 0420 855 509
email: d.oconnor@uq.edu.au website: www.qaafi.uq.edu.au
Online office: https://meet.uq.edu.au/uqdoconn
https://twitter.com/doconnor1874

From: ChipDesign, Madison [mailto:madison.chipdesign@roche.com]
Sent: Saturday, 20 September 2014 12:33 AM
To: Daniel O'Connor
Cc: Agnelo Furtado; RCN Rachaputi; Danoy, Patrick
Subject: Review probe sets for your SeqCap EZ design for Arachis hypogaea (OID 42093, IRN 4000017650, PO 7100020950) ArahyBLNCH

Dear Dr. Daniel O'Connor,

I have completed the Sequence Capture probe selection for your SeqCap EZ design for Archis hypogaea (Roche Nimblegen Order ID 42093, IRN 4000017650, PO # 7100020950).

I have selected two probe sets for your review. The Unique probe set contains up to 3 close matches in the genome as determined by the SSAHA algorithm. We consider a probe to match in the genome if there are five or fewer single-base insertions, deletions or substitutions between the probe and the genome. The Relaxed probe set contains probes with up to 60 close matches in the genome, for the purposes of providing coverage; meaning the vast majority of the probes are unique, with a few probes that have a greater degree of multi-locus homology to increase coverage in all regions.

To assist in your review, I have attached probe coverage summaries. The attached zip archive contains files of the probe positions and the regions that you requested in BED format, veiwable in a number of genome browsers. The file "primary_targets.bed" contains coordinates of your requested regions, which have been modified by padding small targets to a minimum length of 100 bp and by removing overlapping regions. The file “capture_targets.bed” contains the

coordinates of blocks of overlapping probes. The "predicted_no_coverage_regions.bed" contains coordinates of regions from the primary_targets.bed that are not within 100 base pairs of a capture probe.

If you need a BED file viewer to look at the probe locations file, a free version of SignalMap is available at:
http://www.nimblegen.com/products/software/index.html

There are two text files summarizing coverage. The "coverage_summary.txt" describes the global coverage properties of your SeqCap EPI design. This file is tab-delimited text and can be opened in Excel or a text editor. The "coverage.txt" file is tab-delimited text file which contains region-by-region coverage information of the padded and consolidated primary target regions contained in the file "primary_targets.bed". Please see information below for details regarding the contents of each review file.

For a full description of all attached files and the design review process, please see the document at:http://www.nimblegen.com/products/lit/06457312001_NG_SeqCap_EZ_Guide-Review-Approve_v3.pdf

Please note that the review guidance PDF, Step 1.1., mentions a "Design Summary File (PDF)", which is not currently supplied with custom designs; you may safely ignore that step.

Please let me know if you have any questions or would like to make any changes. Once a probe set is approved, I will complete the layout and release the design to manufacturing.

Kind regards,

Nicole

Nicole P. Leahy, Ph.D.
Scientist

Roche NimbleGen, Inc.
Bioinformatics Production
500 South Rosa Road
Madison, WI 53719
USA

Phone: +1-608-316-8013
Fax: +1-608-218-7601
Mailto: nicole.leahy@roche.com
www.nimblegen.com

On Wed, Sep 17, 2014 at 5:34 PM, Daniel O'Connor <d.oconnor@uq.edu.au> wrote:

Great Nicole.

Let me know if there is anything else you need. I will be in early tomorrow morning so should be able to get back to you before 5pm.

Dan

From: ChipDesign, Madison [mailto:madison.chipdesign@roche.com]
Sent: Thursday, 18 September 2014 8:23 AM

To: Daniel O'Connor
Cc: Agnelo Furtado; RCN Rachaputi; Danoy, Patrick
Subject: Re: Question regarding your SeqCap EZ design for Arachis hypogaea (OID 42093, IRN 4000017650, PO 7100020950) ArahyBLNCH

Hello again,

I managed to obtain the file I need. I'll resume work tomorrow morning.

Nicole

On Wed, Sep 17, 2014 at 1:51 PM, Madison ChipDesign <madison.chipdesign@roche.com> wrote:

Hi Dan,

I hit a bit of a snag with the design. In trying to download the genome for Arachis duranensis from peanutbase.org, I received the following error message:

> The requested page "/files/genomes/Arachis_duranensis/assembly/Aradu_v1.0_20140824.fa.gz" could not be found.

Would you know another source for this genome?

Nicole

On Tue, Sep 16, 2014 at 4:53 PM, Daniel O'Connor <d.oconnor@uq.edu.au> wrote:

Great Nicole.

Look forward to seeing what the results are.

Regards,

Dan

From: ChipDesign, Madison [mailto:madison.chipdesign@roche.com]
Sent: Wednesday, 17 September 2014 12:37 AM
To: Daniel O'Connor
Cc: Agnelo Furtado; RCN Rachaputi; Danoy, Patrick
Subject: Re: Question regarding your SeqCap EZ design for Arachis hypogaea (OID

42093, IRN 4000017650, PO 7100020950) ArahyBLNCH

Thanks! I'll start working on these.

On Wed, Sep 10, 2014 at 6:52 PM, Daniel O'Connor <d.oconnor@uq.edu.au> wrote:

Hi Nicole,

Sorry for a couple of days delay getting back to you. I had a few IT issues.

Attached are the gene coordinates for chromosome 1 for both diploids (A genome: duranaensis and B genome: ipanensis).

I think I have them in the required BED format.

Regards,

Dan

From: ChipDesign, Madison [mailto:madison.chipdesign@roche.com]
Sent: Friday, 5 September 2014 1:46 AM
To: Daniel O'Connor; Patrick Danoy
Subject: Question regarding your SeqCap EZ design for Arachis hypogaea (OID 42093, IRN 4000017650, PO 7100020950) ArahyBLNCH

Dear Dr. Daniel O'Connor,

We have received a design specification for your SeqCap EZ design for *Arachis hypogaea* (Roche NimbleGen Order ID # 42093, IRN # 4000017650, PO # 7100020950).

I have a few questions regarding your order. I understand the time difference makes phone calls and Skype difficult to arrange. Hopefully, all questions can be answered through email.

First, from peanutbase.org, there are several files in the assembly directory. My inclination is to use the "by_scaff.tar.gz," but it would be best to use the assembly from which you obtained your coordinates.

Second, the Excel sheet is a bit confusing. We prefer to receive coordinates in BED format (http://bedtools.readthedocs.org/en/latest/content/general-usage.html). This minimizes the risk of introducing errors that may occur by reformatting.

Third, the sheet, Araip.B01_gene model, seems to alternate between mRNA and exons. Assuming "Minimum" is the genomic start position and "Maximum" is the genomic end position, it seems unusual the the mRNAs have so much larger coordinate values compared to the exons. This is not a problem for our pipeline, but it is an oddity that we like to check to make certain their are no errors.

Let me know if you need further clarification of these questions or have any of your own questions.

Kindest regards,

Nicole

Nicole P. Leahy, Ph.D.
Scientist

Roche NimbleGen, Inc.
Bioinformatics Production
500 South Rosa Road
Madison, WI 53719
USA

Phone: +1-608-316-8013
Fax: +1-608-218-7601
Mailto: nicole.leahy@roche.com
www.nimblegen.com

<OID42093_AraHyp_ArahyBLNCH_relaxed_design_deliverables.zip><OID42093_AraHyp_ArahyBLNCH_unique_design_deliverables.zip>