

From: [Scott Allen Jackson](#)
To: [Cannon, Ethalinda K \[E CPE\]](#)
Cc: [Richard Wilson](#); [Lutz Froenicke](#); [Guo: Baozhu](#); [Scheffler: Brian](#); [Corley Holbrook](#); [Soraya Bertoli](#); [david.bertoli@pq.cnpq.br](#); [Shapiro: Howard](#); [Howard Valentine](#); [刘心](#); [Mark Burrows](#); [Peggy Ozias-Akins](#); [rajeev varshney](#); [Ran Hovav](#); [Michelmores: Richard](#); [Steven Cannon](#); [Nwosu: Victor](#); [徐: 琪](#); [Ray Schnell](#)
Subject: Re: peanut assemblies in GenBank
Date: Friday, August 08, 2014 10:41:49 AM

I agree that they should go to GenBank. I think that as the pseudomolecules are in flux, it might be better to submit the scaffolds/contigs. Thoguhts?

scott

On Aug 8, 2014, at 10:32 AM, Cannon, Ethalinda K [E CPE] <ekcannon@iastate.edu> wrote:

Hello PGC members,

In April I recommended that the genome assemblies for *A. duranensis* and *A. ipaensis* be submitted to GenBank, and offered to do the submission. I have just received word that GenBank is ready for the submission. We've had researchers ask why the sequence isn't in GenBank, and it is in general good practice to submit high value sequence to GenBank.

There are two downsides: the process is very slow (multiple months) and the sequence is very likely to change as a result of GenBank's stringent QC pipeline. I'm willing to shepherd the genomes through the process and feel the assembly could be improved by the QC pipeline. We should be prepared to release the GenBank-approved assembly as a new version, perhaps 1.1.

Questions:

1. Are there any objections to submitting the assemblies to GenBank?
2. BioProjects: I'll need to set up BioProject records for the assemblies. I am hoping it will be possible to create a BioProject for each assembly, then create a parent BioProject to link them together. My hope is that this parent BioProject would also link the BioProject for the *A. hypogaea* assembly when that's created. Thoughts on this plan?
3. Submit individual contigs + tiling path or whole psuedomolecules? I have a question in to GenBank to see if it's possible to do both (or advisable), but assuming we have to do one or the other, which should it be?
4. The assemblies should be registered with GOLD (<http://genomesonline.org/>).

This effort is applying MIxS standards (Minimum Information about any(x) Sequence) to genomic assemblies and transcriptomes. This not only provides dependable and standard metadata for the datasets, but GOLD uses the meta data to provide some fairly powerful search capabilities. I am willing to do this too unless there are objections. (There are no Arachis projects listed in GOLD right now.)

Ethy