DRAFT

1st Annual Meeting of the International Cannabis Research Consortium

1pm, Wed Jan 15, 2020, Handlery Hotel, Hotel Circle, San Diego, CA.

MINUTES

Attendees

Graham King - Southern Cross University, Lismore, Australia (*Chair*)
Alisha Holloway - Phylos (*Recorder of Minutes*)
CJ Schwartz - Sunrise Genetics
Jacob Toth - Cornell
Larry Smart - Cornell
George Stack - Cornell
Sophie Watts - Dalhousie - sophie.watts@dal.ca
Hsuan Chen - Oregon CBD - hsuan@jackhempicine.com
Mike Ruckle - Pure Cannabis Research - Switzerland - ruckle@pureeurope.eu
George Weiblen - U MN
Yael Maoz - NRGene - yael.maoz@nrgene.com
Oori Weisshaus - NRGene
Keith Allen - Front Range Biosciences

1. Introductions

It was agreed that the **Mission** of the ICRC should be to:

> *Accelerate genomic, genetic, and breeding research on Cannabis by facilitating collaboration and access to shared resources, and coordinating establishment of standards.*

Goals for 2020 meeting

- Develop community consensus on chromosome ID and numbering scheme
- Discuss potential gene annotation standard vocabulary
- Identify potential areas of collaboration and shared resources
- Determine optimal consortium components - steering committee, communication

2. Updates on status of genome assemblies (NCBI), and 2020 projects.

a. CJ (Sunrise)- resequencing and pan-genome
b. Larry - sequencing a male and female - likely NY feral - working with USDA ARS to establish a hemp germplasm repository.
c. Graham : SCU establishing germplasm resource(s) for hemp and medicinal cannabis - import licenses and facilities in place; Paper on extreme phenotyping GWAS of minor cannabinoids to be submitted by Q2. An AU$10m CRC-P Project on Medicinal Cannabis cultivation (https://www.business.gov.au/Grants-and-Programs/Cooperative-Research-Centres-Projects-CRCP-Grants/CRC-Projects-selection-round-outcomes); A dedicated *Cannabis* bioinformatics PhD student starting Q2.

d. From Noel Cogan (*not present*) @ Ag Vic, Australia
   - For Cannbio-2 (balanced CBD:THC ratio) - transcriptome released 2019; 2020 Q2 release targeted for: PacBio reference genome, aligned to CS10/CBDRx –chromosome nomenclature and orientation completely aligned. Annotated using transcriptome and variety of other tools:
   - More generally: generate wider cohort of short-read resequenced genomes for variants. Focus on medicinal and indoor Ag, so weighted to including high THC strains. Genomic selection and edits.
e. Alisha - genome coming soon
f. George - F1 of THC and CBD cross, lots of Illumina data - can be aligned to reference
g. John McKay (*by email*) - two phased genomes - Carmagnola and USO31 hemp
h. UCDavis - releasing assembly soon – Cherry type
i. Mike - Univ of Stellenbach and another - with OSU - will release, one will be male
j. Kevin (by email) - ??
k. Oori – genomes by NRGene, but can't share sequence

3. Developing a community consensus on chromosome identification and numbering scheme.

The meeting unanimously agreed that a reference genome was required, and now was timely to assign fixed chromosome names/orientation.

- The CS10/CBDRx genome (NCBI GCA_900626175.1) had been annotated by NCBI. Several independent research groups have found this be relatively stable and congruent with other assemblies. Due to the relative quality and completeness, this assembly has by default become considered as a *de facto* reference, although there are a number of acknowledged provisos. These include:
  i. Need to identify the X chromosome
  ii. Completeness of physical coverage telomere to telomore
  iii. Structural variation around THCAS/CBDAS loci etc
- Chromosome numbering an orientation:
  i. John McKay (by email) – had made the case that ideally the chromosomes be ordered by size (large to small). It should be noted that this is by no means universal amongst reference plant genome assemblies.
  ii. However, the meeting agreed it would be worthwhile to allow a short period of time to compare all available high quality genome assemblies and use consensus to number from largest (chr 1) to smallest (chr 9).

- There is the opportunity to incluide the mito and chloroplast genomes from Nolan Kane's lab into the 'mk 1 *Cannabis* reference genome'.

Actions:

- By April 2020, allow time for those who wish and are able - to collate and present evidence of modifications to current CS10 chromosome assignment. Information to be disseminated and assessed by one or more independent lab representatives and agreement established.
- Where and when possible, secure evidence for assignment of X chromosome.
- Final reference should then contain: <chr1…9, X, cp, mt>
- Once agreement is established (preferably by Q3 2020), then disseminate information to research community and key journals. Liaise with NCBI and original submitters to update records;

4. Standardized nomenclature: gene models and functional gene names.

a. **Gene model** nomenclature

i. *Proposal*: as for other genomes, assign unique identifiers to gene models, with consecutive numbering from top to bottom of chromosome, e.g., Cs1g10001.1 (v1 transcript), leaving a gap of 10 integer values to allow for insertions (or more for complex regions of duplication, structural variants). It was accepted that whilst will not be perfect, it will provide a stable framework to move forward.

ii. Assignment should:

1. Be based on the chromosome nomenclature/orientation above
2. First check congruence between available assemblies
3. Identify regions of uncertainty to allow for later input

iii. Todd Michael's dot plots of proteins along chromosome will be a great starting point to guide this process

iv. Keith - White Paper on orthology

b. **Functional** gene names

i. *Proposal*: based on function or homology with symbols for genes italicized (e.g., *CsFLC*) and symbols for proteins not italicized (e.g., CsFLC).

ii. Allow for more detailed discrimination of names, to avoid confusion between paralogs, etc:

1. *Cs1.FLC.a* - *Cannabis sativa*, chromosome 1, *FLC* gene, paralog 'a'

iii. It would be very beneficial to establish a 'registry' to avoid synonyms/homonyms.

It was noted that an Oregon State ontologist would like to get involved.

Actions:

- For gene model naming: Key researchers/groups to carry out check on congruence between available assemblies; make use of Todd Michael's analyses; identify regions of uncertainties.
- Consult between different groups and present proposal for ratification
- Outline process for updated submission to NCBI against reference CS10 and other assemblies.

5. Pan-genome development to incorporate structural variation

a. NRGene has started down this path - can include additional genomes, but funding is an issue
b. The meeting agreed that to date there was currently insufficient data available.

Action: Review in 2021

6. Public Dissemination

a. Once a reference genome assembly, set of gene models, and nomenclature have been established and agreed, update at NCBI. This then opens opportunity to have *Cannabis* genome curated within **Ensembl** Plants. The EBI team require evidence that is representative of a community/consortium accepted effort. Graham can help in liaison with the Ensembl Plants group.

b. **Funding** for community efforts – require evidence of common vision and set of goals in order to draw down national public sector funding, and will help leverage other funds.

c. Scope for publishing a **White Paper**, to outline current position, scope, opportunities and unique challenges for *Cannabis* genomics:
    i. Able to outline progress to date, announcing chromosome numbering, annotation conventions
    ii. Issues of non-autosomal chromosomes
    iii. Validation of germplasm
    iv. Complexity of describing legal status (eg to lawyers) and how genomics can assist.

d. **Web site** – it was agreed that there would be benefits from establish a point of reference and record that is clearly arising from peer-reviewed scholarly research. There is a need to establish:
    i. Editorial overview and responsibility
    ii. Hosting that is independent of perceived commercial conflicts of interest
    iii. Would an academic institution be willing to host/contribute to editing?
        1. Small resource input required for site design/set up
        2. Time for editing and content management

Actions:

- Identify who willing to host and procure site design/set up
- Identify editorial overview/site content update responsibilities

7. Sharing germplasm

a. There is an ongoing requirement for conversation about how to share, what is required

Action:

Assign this to agenda of next meeting

8. Conclusions and next steps

a. Naming of the Consortium – there was discussion on scope, and whether the title should include the term 'genome' - e.g. **International *Cannabis* Genome Research Consortium** (ICGRC), to retain the focus on key aspects of genomics, genetics and genetic resources.

Action: to be decided

b. It was agreed that it would be beneficial during 2020 to have quarterly conference calls – next one in April (tbc)

Actions:

- date and time to be set by Alisha (?)

- send email of meeting minutes, first as draft to those attending, and then to wider group including those unable to attend, and ask those to add other key contributors to field, so a global list of relevant interested parties are included.

Original Agenda

1. Update status of genome assemblies (NCBI), and 2020 projects. (10min)
2. Develop community consensus on chromosome identification and numbering scheme (15 min)
3. Present gene annotation standard vocabulary ideas. (15 min)
    a. Gene model nomenclature
        i. Proposal: unique identifier with consecutive numbering from p to q on chromosome, e.g., Cs1g10001.1 (v1 transcript)
    b. Functional gene names
        i. Proposal: based on function or homology with symbols for genes italicized (e.g., *CsFLTC*) and symbols for proteins not italicized (e.g., CsFLTC)
        ii. Establish a 'registry' to avoid synonyms/homonyms.
4. Discuss moving towards a pangenome to incorporate structural variation (20 min)
5. Cannabinoid synthase cassettes
    a. Huge variation in content, order and orientation.
    b. Discuss thoughts on incorporating synthase cassettes into assemblies.
6. Shared resources
    a. Open register of experimental germplasm
    b. Cannabis Genomic database
        i. Get 'reference' genome in queue for Ensembl which allows for a wide range of tracks to be added.
        ii. Brief update on EBI developing pipelines to be more plant/polyploid relevant.
    c. Germplasm bank (tissue culture)
    d. Seed bank
        i. Regulatory issues for high THC plant seeds
        ii. Hemp seed bank at Cornell
7. ICRC Functions and Processes - discussion led by Graham King
8. Steering Committee & Membership
    a. Number of members? 5-9, with one chair
    b. Other roles? Secretary for note taking and/or collation.
    c. Term of office?
    d. Majority of Steering Committee has university/government/non-profit appointments.
    e. Membership open to all with emphasis on diversity and inclusion.
9. Communication
    a. Annual meeting at PAG.
    b. Release summary statement to all ICRC members following annual meeting.
    c. Email newsletter? Frequency?
    d. Web site? Would allow for membership signup.
    e. External communication to scientific community - white paper?
    f. Ensure that we reach out to others in Cannabis Genetics, Genomics, and Breeding who may not be at the meeting.
        i. John McPartland (University of Vermont, GW Pharma)
        ii. Jon Page (UBC, Aurora Cannabis)

iii. Nolan Kane (University of Colorado)
iv. Dario Cantu (UC Davis)

10. Funding
    a. How to obtain funding to support yearly meetings?

- Potential founding Steering Committees:
    - John McKay (Colorado State University, New West Genetics)
    - George Weiblen (University of Minnesota, Sunrise Genetics)
    - Graham King (Southern Cross University)
    - Sean Myles (Dalhousie University)
    - Larry Smart (Cornell University)
    - Alisha Holloway (UCSF, Phylos)
    - Noel Cogan (LaTrobe University, AgriBio)
    - CJ Schwartz (Sunrise Genetics)
    - Kevin McKernon (Medicinal Genomics)