MBGP Agenda PAG 2023

- Approval of minutes from previous meeting
  Rod and Graham approved

- Round-table reports on progress/notable milestones/announcements for members

RS – project running to sequence 100 napu with long reads (nanopore), Rod give list of genotypes to stop overlaps. Data will be available. From ASSYST, range of different genotypes. No data analysis. Phenotypic data also. 40 x not going to assemble. Mapping to armor 10 and express. Look at SV. Also running pacbio on 6 accessions. Re-arrangements between generations

IP – sequenced 50 from founders of NMA population., ONT, illumine polishing, SV calling. Data available when can. All springs, some semi-winter.

Ian – re-seqeunced radiation (gamma, fast neutron) mutation population, available for collaboration. Can test for impact of CNV. 800 re-seq (12x) have 2000. (European winter). No database for mutations, have scoring matrix.

Rod – TILLING population for express. Re-seq with illumine, should be available, is website. Look for mutations. Owned by NPZ, can contact, should be open to releasing them. 3000 lines, not sure how many re-seq.

Cyril- 5 diploid assemblies with genoscope, 2 A and 3 C ONT 50x, not anchoring with genetic maps.

IB – happy to check organisations

AG. Just re-assembled express with hi-fi

Mia – about to submit assembly for tournforitei, poacbio, nanopore hi-ci. Doesn't like A, look into more. 212.

DE – master of bioinformatics course, as part of course there are projects. Looking for partners and collaborations.

Can assemble after publishing -Yin.

- Mutant sequencing resource (Christian Jung)
- Diploid Pan genome annotation (please see the paper sent out by Graham)

Previous meetings, principle to agree to adopt annotation, A01xx etc denote species and genotype. Huge amount work by Ian, establishing diploid based pan genome (He et al 2021), ABC genes, one initial genotype seen as high quality. Any new genotype, annotated genes

follow same form. Implement into modern and existing assemblies, need to go back and re-label existing annotated genes. Ro18.

Ensembl plants – only one that is there with genome nomenclature. Table of existing genomes, assembly names, indication of when went in, which used for genome and pan genome paper. List ensure annotated and labelled, so consistent.

IP – clarify original nomenncalture and synonym label, haven't thrown away genes.

What needs to be done for other genoems. Person responsible, only people allowed to change or give permission. Identify who they are, generate gff to submit to NCBI. Without that, with community agreement, can look

Sarah. Can have synonym as priority if want.

Dave – is there automated method to generate synonyms. Graham had someone, but gone back, might be able to pass on. Dave to talk to Graham afterwards.

IP – what about updating pan genome.

GK – paper is core, gap of 10 genes between, script would need to add in P for synonym.

GK- what is criteria for allele?

IB – aimed at ancestral arrangement.

DE- will be resolved in graph

IB – most are just genes not originally called.

JB – need to get permission from the owners. Need original gff, must be submitted to ncbi

AG – need to ensure annotation from author, not NCBI.

ACTION – owners will be contacted. 2 nigra genomes JB work with Graham on this. Chromosome number confusion on B genome. Some inverted. Re-order first.

Sarah – the assemblies need to be added to ensemble. No Brassica funding, could have prioritistaion for community, will add in.

AG – has github how to submit gff, can send on. Can circulate. AG happy to help.

- International collaborative projects

IB – working with Jun Zhou on carinata assemblies on new genomes and pan genome. Eliminated errors. Taking public to make pan genome. New are better organized compared to 2 published. IP has one she can also provide.

AH – proposal – BBSRC/NSF project. Brassicas model. Is about tools and infrastructure – ways of building and viewing pan genomes. Eric Garrison (Tenessee) -developed PGBB, moving from human to plants. Needs optimizing to plant genomes. Identified variation not seen. Need to develop tools, not optimized, Ensembl, paying for someone in Sarah's group. Upload genomes into Ensembl. Will generate pan graphs, intra not inter species. Store pan graphs. Linear graphs, identify loci interested in. Get sub graph of how varies. Tool anyone can use. Move to look at other species. Map short reads back to graph, rather than reference bias, GWAS – against graph, PWAS. Just submitted.

Comparison to other tools?

Graham- how will it look in ensembl? Linear genomes, look as everything else. Portal that can drop pan graph. Using ensemble as portal. Pan graph meta data standards.

Start a.t, move to rapa, move to napus. Only use genomes in best shape. Also workshops, would meet with us, discuss genomes, workshops at each stage, come and understand how work. Engagement with pan graph. Would start Sept if funded and 4 year grant. Yr 1, at, 9 months after rapa.

GK – checking if re-annotate, No, so need to make sure add labels of pan genomes.

GK – how do we encourage people to add at source before publish.

Reviewers, editors?

GK - Criteria for match needs to be made clear to people. He is happy to help
AG – need tool to check, to make it easy.

DE- graph pan genomes, wants data for this,

GK – is Ian's paper core, or does it re-rrange.
DE – is new graph, as not linear

Sarah – google summer of code student.

- Brassica 60K/90K array replacement
- Meeting announcements

Rapeseed congress

Crucifer Genetics – what year?

ISHS postponed – India, looking for another country – maybe 2024? Maybe Korea combine the 2?

- Any other items